

ADVANCED BOUNDARY CONDITION METHOD IN QUANTUM TRANSPORT
AND ITS APPLICATION IN NANODEVICES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yu He

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Professor G. Klimeck, for the valuable insights he provided for this research and the continuous support of my Ph.D. I would like to thank him for providing all the necessary computing resources. He introduced me to the world of high performance computing and its applications to the problem solving in semiconductor industry which will be invaluable experience for my future career pursuits.

I would also like to thank Professor S. Datta, Professor L. Rokhinson, and Professor T. Kubis for serving on my advisory committee.

I am grateful to Professor T. Kubis provided a lot of supports during the past few years of my Ph.D. He provided great help to improve the quality of my work. He shared with me his knowledge in device physics and quantum transport which are very insightful for my work. I would also like to thank Professor M. Povolotskyi for the discussions and supports in software development.

I would also like to thank our research team of Dr. Jim Fonseca, Dr. Bozidar Novakovic, Dr. Arvind Ajoy, and Dr. Jun Huang who have been a great support.

I would also like to thank our wonderful group members and friends Yaohua Tan, Junzhe Geng, Zhengping Jiang, Kai Miao, Yuling Hsue, Yu Wang, Pengyu Long, Fan Chen, Dr. Xufeng Wang and Dr. Lang Zeng. We have spent many happy hours together for interesting discussions and many events.

I am grateful to Daniel Mejia for the technique support he usually provided. I would like to thank James Charles, Hesameddin Ilatikhameneh, Daniel Lemus, Prasad Sarangapani, Dr. Ganesh Hegde, Dr. Sung Geun Kim, and Dr. Seung Hyun Park for interesting and helpful discussions.

I would like to thank all my colleagues and administrative support from the NCN and the Klimeck group for providing a stimulating and fun environment in which to learn and grow.

I am grateful to my wife Biying Liu, with her life in campus becomes colorful. Lastly, and most importantly, I wish to thank my parents and family for their love and support.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	ix
ABSTRACT	xv
1. INTRODUCTION	1
1.1. Objectives.....	1
1.2. Summary of This Work	3
2. QUANTUM TRANSPORT IN EMPIRICAL TIGHT BINDING.....	5
2.1. Introduction	5
2.2. Quantum Transport Methods	6
2.2.1. NEGF and RGF	6
2.2.2. QTBM.....	8
2.2.3. Boundary condition: contact self-energy	9
2.3. Empirical Tight Binding Method	12
2.3.1. Motivation.....	12
2.3.2. Tight binding Hamiltonian: formula and notation	13
2.3.3. Brief introduction to the TB mapping method.....	14
2.4. Summary	16
3. ALGORITHM IMPROVEMENT AND IMPLEMENTATION	17
3.1. Introduction	17
3.2. RGF Implementation	18
3.3. QTBM Implementation.....	20
3.4. Implementation of Transfer Matrix Method	22
3.5. Summary	24
4. LOW RANK APPROXIMATION METHOD	25
4.1. Introduction	25
4.2. LRA in Effective Mass Approximation.....	26
4.2.1. Reduced real space basis.....	26
4.2.2. Comparison with existing methods	29
4.2.3. Numerical complexity and memory usage analysis	30
4.2.4. Benchmark and application.....	32
4.2.5. Summary.....	40

	Page
4.3. LRA in Tight Binding.....	41
4.3.1. Basis function.....	41
4.3.2. Benchmark and application.....	42
4.3.3. Common basis in LRA-TB	48
4.3.4. Summary.....	51
4.4. LRA in Contact Self-energy	51
4.5. The Eigenvalue Solvers	55
4.6. Summary.....	55
5. GENERAL LEAD ALGORITHM FOR CONTACT SELF-ENERGY	57
5.1. Introduction.....	57
5.2. Algorithm Details	58
5.2.1. The contact structure	58
5.2.2. The iterative solution.....	58
5.2.3. Simple scattering model in contact	60
5.3. Benchmark and Application.....	62
5.3.1. Si nanowire with periodic contact.....	62
5.3.2. Graphene nanoribbon with trumpet shape contact.....	64
5.3.3. Alloy disorder in contact.....	66
5.4. Summary.....	69
6. SURFACE PASSIVATION METHOD.....	70
6.1. Introduction.....	70
6.2. Method Details	71
6.2.1. Self-energy of the passivation atoms.....	71
6.2.2. Parameterization.....	72
6.3. Benchmark and Application.....	73
6.3.1. Benchmark with known approach.....	73
6.3.2. Application on Si/SiO ₂ interface	73
6.4. Summary.....	80
7. BLACK PHOSPHORUS TRANSISTOR	81
7.1. Introduction.....	81
7.2. Method.....	82
7.2.1. TB parameterization	82
7.2.2. Acoustic phonon scattering.....	84
7.2.3. Surface roughness scattering.....	85
7.2.4. Mobility calculation.....	86
7.3. Results and Discussions.....	86
7.3.1. Band structure	86
7.3.2. Mobility	90
7.3.3. ITRS 2020 target transistor.....	93
7.4. Summary.....	94

	Page
8.CONCLUSION AND OUTLOOK	95
LIST OF REFERENCES.....	98
APPENDICES	
A. Revisit The LRA Basis for Effective Mass.....	106
B. Roughness Implementation.....	107
VITA.....	111
PUBLICATIONS	112

LIST OF TABLES

Table	Page
2.1 Notation of TB parameters.....	14
3.1 Peak memory usage for RGF in Gigabytes. std, the state of literature, opt, the optimized way in this work. It shows that the peak memory does not increase significantly with more energy points with the stated optimizations.....	20
3.2 Timing comparison in seconds, for 1 energy point. std, the state of literature, opt, the optimized way discussed in this work. It shows that the stated optimizations lead to a total speed up of 5x in the contact self-energy calculations and 35x in the construction of the QTBM RHS matrix.	24
4.1 The measured time consumptions of ballistic transport calculations for a 100nm long GaAs resistor with different methods. The results are solved in MATLAB in serial. Speed ups are observed in all the established efficient NEGF algorithms. The LRA method shows a similar performance compared to the CBR method in this example.....	30
4.2 The measured time consumptions for the three examples with matrix rank reductions down to 10% and 20% as well as the exact NEGF solutions are listed. The LRA method and the standard NEGF calculations were implemented in MATLAB with 8 cores parallelization. The results show significant speed ups of the LRA method.....	36
5.1 The measured time and memory comparisons between different contact self-energy methods: Sancho-Rubio, transfer matrix, and general lead. The calculations are done for the same Si nanowire as in Fig. 5.4, with 1 energy point only. The general lead method shows the lowest memory usage, and 6x faster than the Sancho-Rubio method.....	64
6.1 Passivation parameters in units of eV.....	75
6.2 Tight binding (TB) and <i>ab-initio</i> (DFT) band gaps E_g , valence E_v and conduction E_c band edges in eV, and effective masses for electrons m_e and holes m_h at the Γ point along X and M directions for the oxidation configurations of Fig. 6.2. The TB results match well with the DFT calculations.....	76

Table	Page
7.1 Tight binding parameters for black phosphorus in eV. In-plane parameters are denoted by α , and inter-layer parameters are denoted by β	87
7.2 Tight binding (TB) and <i>ab-initio</i> (DFT) band gaps E_g , valence E_v and conduction E_c band edges in eV, and effective masses for electrons m_e and holes m_h at the Γ point along X and Y directions for the monolayer, bilayer, and bulk black phosphorus. The TB results match well with the DFT calculations.	88

LIST OF FIGURES

Figure	Page
1.1 (a) Square $\text{Ga}_{0.5}\text{In}_{0.5}\text{As}$ random alloy nanowire with diameter of 4nm and device length of 6nm and 18nm. (b) Lower transmission for the longer device indicates that fewer electrons can tunnel through the device, which is attributed to the DOS mismatch between the pure contacts and the alloyed device. VCA result is shown for comparison.....	2
1.2 (a) Band structures for a 2.2nm thick Si quantum well with (red) and without (black) passivation show that passivation removes the surface states in the band gap. The wave function of the surface state is shown in the inset. (b) Different passivation potentials lead to different band structures for the same quantum well suggesting that passivation can introduce artifact in the volume states.....	3
2.1 A schematic of a typical device in transport modeling is shown. The device is gated in the central region and attached to two contacts at both ends.	5
2.2 The idea of iterative solutions in the Sancho Rubio method suggests an exponential extension of contact blocks which ensures fast convergence.	10
2.3 The process of mapping <i>ab-initio</i> results to TB.....	15
3.1 RGF process for (a) forward iterations and (b) backward iterations. Yellow blocks denote the matrix elements that are stored. The yellow lines in (b) indicate that only the diagonal elements are stored during the backward iterations.....	19
3.2 Speed up factor of NEMO5 QTBM calculations with respect to number of cores used, for 20nm long Si nanowires with different diameters. It suggests that the calculations are close to ideal scaling performance up to 4 cores.....	21
4.1 Spatially resolved current density in the homogeneous structure with a linear potential drop of 0.2eV. The lines show that the result calculated approximately with a reduction of the matrix rank down to 20% (dashed) agree well with the NEGF method solved exactly (solid), and the result with a reduction of the matrix rank down to 10% (dash-dotted) shows a larger deviation from the exact solution.....	33

Figure	Page
4.2 Calculated electron density of the homogeneous device of Fig. 4.1 in (a) equilibrium and (b) when a linear potential drop of 0.2eV is assumed. Results with approximated NEGF calculations where the matrix rank is reduced down to 20% (dashed) agree well with the exact NEGF calculations (solid). Results with matrix rank reduction down to 10% (dash-dotted) show larger deviations from the exact solution.	34
4.3 Comparison of I - V characteristic of the structure of Fig. 4.1 shows that the approximated NEGF solutions where the NEGF matrix rank is reduced to 20% (dashed) agree well with the exact NEGF calculations (solid). The results with matrix reduction down to 10% (dash-dotted) show larger deviations. Results for the NEGF calculations with a 10 times coarser real space mesh (dotted) are shown to deviate significantly from the dense mesh case due to its different effective electron dispersion.	35
4.4 Conduction band profile (solid line) and contour plot of the energy resolved electron density of the RTD structure described in the main text. The energy resolved density is calculated within the NEGF method (a) exactly and (b) approximately by a reduction of the NEGF equations' matrix rank to 10%. The results show that the filling of bound states in the triangular well by inelastic acoustic phonon scattering is well captured in the approximated method.	37
4.5 I - V characteristic of the RTD structure of Fig. 4.4 calculated approximately with 10% (dashed) of the original matrix rank agrees well with the exact solution (solid). However, too aggressive matrix rank reduction of 3.1% (dash-dotted) fails to reproduce the results due to the neglecting of relevant electronic states.	38
4.6 The I - V characteristic of the 1000nm homogeneous structure described in the main text when different values for the deformation potential are used: 27eV (dashed), 60eV (dash-dotted), and 135eV (dotted), and ballistic results are shown as a solid curve. All results are determined from approximated NEGF equations with a matrix rank reduction down to 20% of the original rank. With the deformation potential of 135eV the I - V characteristic is almost ohmic.	39
4.7 Energy resolved electron density of the 1000nm long structure of Fig. 4.6 with a deformation potential of 135eV and a potential drop of 0.1V. Energy dissipation happens around the center of the device suggesting a transition from effectively ballistic transport into the drift-diffusion of the rightmost 500nm of the device.	40
4.8 Relative deviation of electron density in a squared, 5nm diameter Si nanowire of exact NEGF calculation and of NEGF calculations approximated with 10% of the original matrix rank. Transport is along the x-axis. The small relative deviations suggest that with 10% of the original matrix rank the electron density is nicely reproduced.	43

Figure	Page
4.9 Electron density along the center of the nanowire shown in Fig. 4.8, of the exact NEGF calculation (circle) and of NEGF calculations with 10% of the original matrix rank. The small deviations suggest that with 10% of the original matrix rank the electron density is well reproduced.....	44
4.10 Total transmissions of the exact NEGF calculation (solid) and of NEGF calculations with reduced matrix rank for the valence band of the Si nanowire shown in Fig. 4.8. With 10% of the original matrix rank the transmission is well reproduced, while too aggressive reduction of 2% of the original rank leads to larger deviations.	45
4.11 Total transmissions of the exact NEGF calculation (solid) and of NEGF calculations with reduced matrix rank for the conduction band of the Si nanowire shown in Fig. 4.8. With 10% of the original matrix rank the transmission is well reproduced, while too aggressive reduction of 2% of the original rank leads to larger deviations.....	45
4.12 Total transmission of a 3nm diameter, 8nm long Si nanowire with 0.1eV linear potential drop is solved. It shows that NEGF calculations with 10% of the original matrix rank (black) nicely reproduce the transmission of the exact NEGF calculation (red).....	46
4.13 L-shaped GaSb-InAs TFET transistor.	47
4.14 The total transmission for the GaSb-InAs TFET shown in Fig. 4.13, with exact NEGF calculation (red) and approximated NEGF calculation of 5% of the original matrix rank (black). The band edges of the GaSb and InAs are shown to indicate the energy range for tunneling. The results are well reproduced with LRA in both valence band, conduction band, and the broken gap range where tunneling happens.	47
4.15 Total transmission calculated with 10% of the original matrix rank for the valence band of a 12nm diameter Si nanowire without a supercomputer. The inset shows the schematic of the nanowire. This calculation shows the power of the LRA method.	48
4.16 Total transmission calculated with 10% of the original matrix rank for the conduction band of a 12nm diameter Si nanowire without a supercomputer. This calculation shows the power of the LRA method.	48
4.17 Transmissions of the valence band and conduction band for a 12nm long quasi-1D GaAs structure. Calculations are done exactly and with LRA of 10% of the original rank. The basis functions are generated in a few specific energy points only. In the LRA case, the transmission is reproduced only in a tiny interval of the energy points where the basis functions are generated.....	49

Figure	Page
4.18 Transmissions of the valence band and conduction band for a 12nm long quasi-1D GaAs structure. Calculations with LRA of 10% of the original matrix rank agree well with the exact calculations. The basis sets are generated following (4.13). The results suggest that the new method of (4.13) is a potential solution in finding a common basis.	51
4.19 Typical eigenvalue spectrum of (2.18). The propagating modes are corresponding to the real part of -0.5eV , and the decaying modes are decaying faster away from that.	53
4.20 Transmission solved with LRA of 10% of the original rank matches well with the exact solution.	54
4.21 Density solved with LRA of 10% of the original rank matches well with the exact solution.	54
5.1 Schematic of the general lead algorithm program flow. The algorithm iterates from index 0 to N . The numerical contact portion of the algorithm has a length of N slabs.	60
5.2 (a) Conduction band profile for an $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}/\text{GaAs}$ RTD. The device is divided into three regions: left/right reservoirs and the central device. The reservoir regions are in thermal equilibrium. (b) The spectral function close to the central device region shows that the quasi-bound states are present and inelastic scattering is required to feed these states.	61
5.3 Spectrum for Si UTB with in-plane momentum at 0.15BZ. Left: with $i^*\sigma$ only above the bandedge, the band tails do not exist; Right: with $i^*\sigma$ present in the bandgap, the band tails are clearly seen close to the boundaries.	62
5.4 Transmission comparison for a Si nanowire shows that the case with contact self-energy solved with general lead method (symbols) agrees well with the case with contact self-energy solved with transfer matrix method (solid line).	63
5.5 I_d-V_g curve for a 20 nanometer long $n-i-n$ Si nanowire FET with $V_{ds}=0.3\text{V}$. The high doping density is 10^{20} cm^{-3} . The intrinsic gated region is 8nm long with oxide thickness 1nm. It shows that with the self-energy solved with general lead method (black line) the results agree well with the case when the self-energy is solved with transfer matrix method (red line).	64
5.6 (a) Graphene nanoribbon with trumpet shaped contacts: device 1 is defined only in the central square region; device 2 is the central region plus an extension of 5nm into the trumpet shape region. The contacts are the rest of the trumpet shape GNR. (b) The transmission for device 1 (dashed line) agrees well with device 2 (circle) showing that the general lead method works well in the trumpet shaped contacts.	65

Figure	Page
5.7 (a) Schematic of randomly alloyed $\text{Si}_{10.5}\text{Ge}_{0.5}$ nanowires with pure Si contacts and contacts with $\text{Si}_{10.5}\text{Ge}_{0.5}$ random alloy. (b) Calculated transmissions for the alloy randomness structures as shown in (a), with device lengths of 20nm and 6nm, respectively. In the alloyed contact cases the transmissions show length independent properties suggesting that the DOS of the contacts matches well with the device. The transmission in VCA is shown for comparison.	67
5.8 I_d-V_g for the three devices shown in Fig. 5.7(a) and the VCA result. All current densities are solved with $V_{ds}=0.3\text{V}$. The random distribution of alloy atoms in the contacts reduces the ON-state current density by 42% compared to ideal contacts which suggests that having non-trivial contacts is critical in quantitative prediction of device performance.	69
6.1 Bandstructures for 3nm diameter nanowires with (a) Si and (b) GaAs, calculated with the published method (red line) and the new method (black symbol). The results show that the present method can reproduce the published method with proper parameters.	73
6.2 (a) Atomic structure of the three Si/SiO ₂ configurations described in the main text. Spheres represent Si (dark gray), oxygen atoms with different surroundings O1, O2, and O3 (black), and H (light gray). The dispersion relations of <i>ab-initio</i> (solid) and TB (dashed) calculations of a 2.2nm thick Si quantum well oxidized in (b) DBM, (c) BOM, and (d) HGM configurations agree well.	74
6.3 <i>Ab-initio</i> wave functions for (a) DBM, (b) BOM, and (c) HGM oxidation configurations at Γ points. Left: for conduction band, right: for valence band. The results suggest that in the DBM and BOM configurations the surface states are not totally suppressed, while in the HGM case the surface states are well suppressed.	77
6.4 Density of states comparisons of a 2.2nm Si quantum well for (a) HGM vs implicit H passivation, and (b) the three Si/SiO ₂ configurations of Fig. 6.2. The results suggest that the DOS is higher with more O atom contributions in the system.	78
6.5 Calculated optical band gaps of Si (100) quantum wells with varying thicknesses oxidized in the three configurations of Fig. 6.2. Experimental data of [70] (asterisks) are given for comparison. The HGM configuration produces optical band gaps in good agreement with experiment.	79
6.6 I_d-V_g characteristic of the 3.3nm Si UTB transistor described in the main text when dangling bonds are passivated implicitly with H atoms following [18] (dashed) and with the present method for the HGM (solid) oxidization configuration. The results suggest that passivation with only H atoms overestimates the transistor performance.	80

Figure	Page
7.1 Black phosphorus structure in (a) top view and (b) side view. The Brillouin zone for (c) monolayer and (d) bulk structures.	84
7.2 Band structures for (a) monolayer, (b) bilayer, and (c) bulk black phosphorus solved with TB (black dashed line) agree well with <i>ab-initio</i> (red solid line) calculations.	89
7.3 (a) Interface roughness in a four-layer (1.8nm) black phosphorus, and (b) band structures with (black solid line) and without (red dashed line) passivation. It shows that with passivation the surface states are removed from the band gap.	90
7.4 Calculated hole mobility considering (a) acoustic phonon scattering and (b) roughness scattering for transport along armchair (red circle) and zigzag (black triangle) directions. Experimental data (blue star) from [87] is shown for comparison. The results show that the phonon assisted mobilities are much higher than measurement which suggests that acoustic phonon is not the dominating mechanism in the transport properties of black phosphorus. The mobilities of surface roughness agree well with measurement suggesting that roughness scattering might be the dominating mechanism.	92
7.5 (a) Structure with roughness (b) I_d-V_g characteristic for 3.4nm thick black phosphorus UTB transistors along armchair (red circle) and zigzag (black triangle) directions. The results show that the transport properties of the armchair direction outperform the zigzag direction, supporting the experimentally observed anisotropic transport properties.	93
 Appendix Figure	
A.1 Surface roughness with (a) $L_c=0.2\text{nm}$, (b) $L_c=0.7\text{nm}$, and (c) $L_c=1.4\text{nm}$ show the right trend where larger correlation length produces larger rough “islands”.	109
A.2 (a) $I-V$ characteristic for a 3nm diameter, 20nm long Si nanowire in ballistic and with roughness scattering (b) the current density ratio between rough case and ballistic case. It shows that roughness scattering reduces the ON-state current density by 40% compared to ballistic calculations.	110

ABSTRACT

He, Yu Ph.D., Purdue University, December 2015. Advanced Boundary Condition Method in Quantum Transport and Its Application in Nanodevices. Major Professor: Gerhard Klimeck.

Modern semiconductor devices have reached critical dimensions in the sub-20nm range. During the last decade, quantum transport methods have become the standard approaches to model nanoscale devices. In quantum transport methods, Schrödinger equations are solved in the critical device channel with the contacts served as the open boundary conditions. Proper and efficient treatments of these boundary conditions are essential to provide accurate prediction of device performance. The open boundary conditions, which represent charge injection and extraction effects, are described by contact self-energies. All existing contact self-energy methods assume periodic and semi-infinite contacts, which are in stark contrast to realistic devices where the contacts often have complicated geometries or imperfections. On the other hand, confined structures such as quantum dots, nanowires, and ultra-thin bodies play an important role in nanodevice designs. In the tight binding models of these confined structures, the surfaces require appropriate boundary treatments to remove the dangling bonds. The existing boundary treatments fall into two categories. One is to explicitly include the passivation atoms in the device. This is limited to passivation with atoms and small molecules due to the increasing rank of the Hamiltonian. The other is to implicitly incorporate passivation by altering the orbital energies of the dangling bonds with a passivation potential. This method only works for certain crystal structures and symmetries, and fails to distinguish different passivation scenarios, such as hydrogen and oxygen passivation.

In this work, an efficient self-energy method applicable for arbitrary contact structures is developed. This method is based on an iterative algorithm which considers the explicit contact segments. The method is demonstrated on a graphene nanoribbon structure with trumpet shape contacts and a $\text{Si}_{0.5}\text{Ge}_{0.5}$ nanowire transistor with alloy disorder contacts. Furthermore, a new surface passivation method for passivation of arbitrary crystal structures and symmetries is developed. This method is based on a self-energy treatment of the passivation materials and the parameters are validated with *ab-initio* calculations therefore it can distinguish different passivation scenarios. The method is illustrated on the Si/SiO_2 interface with different oxidation configurations. These methods are applied to study the surface roughness and its impact in the transport properties of black phosphorus transistors. The calculated mobilities due to roughness scattering agree well with measurement suggesting that roughness might be the essential scattering mechanism in black phosphorus transistors. The calculated transport properties along armchair direction outperform the zigzag direction, further demonstrating the experimentally observed anisotropic behavior.

1. INTRODUCTION

1.1. Objectives

With the continuous development of the semiconductor industry, the dimension of electronic devices is shrinking pretty fast. As the dimension of devices reach sub-20 nanometer scale, quantum effects such as confinement, tunneling and interference become crucial in the device performance. Classical transport approach based on drift-diffusion equation cannot capture these quantum effects accurately. During the last decade, quantum transport methods such as Non-equilibrium Green's Function (NEGF) [1]-[5] and Quantum Transmitting Boundary Method (QTBM) [6]-[9] have consistently shown to provide a reliable description of quantum effects in nanodevices.

Despite the many advantages of quantum transport methods, several challenges exist. The first critical challenge is the efficiency. As an example, to solve a ballistic I - V characteristic with 10 bias points for a 3nm diameter 20nm long Si nanowire field effect transistor (FET) in NEGF requires $\sim 100\ 000$ seconds. If phonon scattering is taken into account, this can be another factor of $\sim 100x$ slower. This is way below the classical transport approach which typically takes only a few seconds. The second challenge relates to the approximations made in the boundary condition. In order to solve quantum transport efficiently, the Schrödinger equation is solved only in the critical device channel while the contacts are the boundary conditions. The charge injection and extraction of the contacts are solved with contact self-energies. In all the existing self-energy methods, such as the Sancho-Rubio method [10] and the transfer matrix method [6], a periodic and semi-infinite contact structure is assumed, which is in stark contrast to realistic devices whose contacts are usually not periodic and have imperfections like alloy randomness or defects. Such a contradiction introduces mismatch of density of states (DOS) in the device/contact interface which results in the strong artificial reflections of

carriers and affects the transport results significantly. As an example, Fig. 1.1 shows the transmission comparison for a $\text{Ga}_{0.5}\text{In}_{0.5}\text{As}$ random alloy nanowire with 18nm and 6nm length. In such randomly alloyed device simulation, it is commonly seen in literatures to use pure materials in the contacts as required by the semi-infinite assumption [11]-[13]. In Fig. 1.1 pure InAs wire is taken as the contact material. The transmission of the 6nm device is much larger than that of the 18nm case. This indicates the tunneling of pure InAs contact electrons through the random alloy $\text{Ga}_{0.5}\text{In}_{0.5}\text{As}$ device: the longer the device, the less likely the tunneling. This tunneling effect is present due to the DOS mismatch of the pure InAs contact and the alloyed device, which is an artifact since the major physics is not only dominated by the device itself, but also the contact (environment). The transmission result of virtual crystal approximation (VCA) [14] is also shown for comparison.

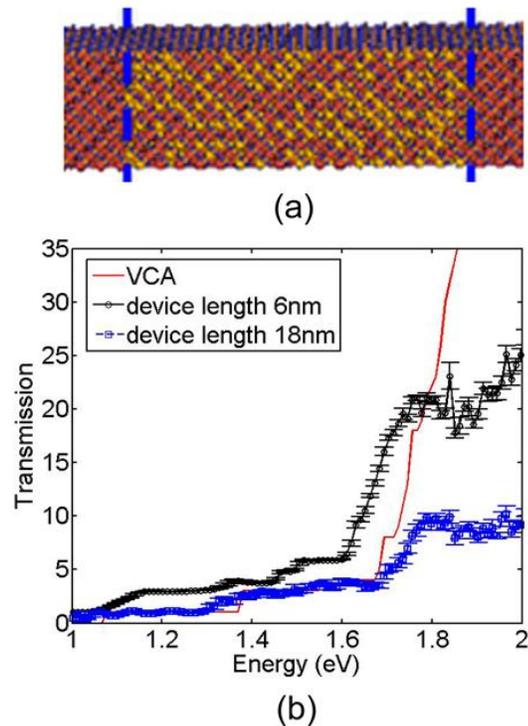


Fig. 1.1 (a) Square $\text{Ga}_{0.5}\text{In}_{0.5}\text{As}$ random alloy nanowire with diameter of 4nm and device length of 6nm and 18nm. (b) Lower transmission for the longer device indicates that fewer electrons can tunnel through the device, which is attributed to the DOS mismatch between the pure contacts and the alloyed device. VCA result is shown for comparison.

Another challenge is in the treatment of the device surfaces. Surface atoms contribute dangling bonds which result in surface states present in the energy gap of the material. Experimentally these surfaces are passivated with passivation atoms, such as hydrogen or oxygen, or covered with an oxide [15]-[17]. In simulation, however, explicit inclusion of passivation atoms or oxide can be numerically very expensive since it increases the matrix rank of the Hamiltonian significantly, especially for nanodevices with a large surface-to volume ratio. Alternatively, an implicit passivation treatment is usually adopted. In this implicit treatment, the orbital energies of the dangling bonds are altered with a passivation potential [18]. With the proper passivation potential the surface states will be removed from the critical energy range where transport usually happens. Fig. 1.2(a) shows the band structures for a 2.2nm thick Si quantum well with and without passivation. In the case without passivation the surface states are clearly seen in the band gap of the quantum well. The wave function of a surface state shown in the inset indicates that this state exists on the atom that is not passivated. The red lines in Fig. 1.2(a) shows that with a passivation potential added to the surface atoms the surface states are eliminated. However, Fig. 1.2(b) shows another problem of this passivation scheme: two different passivation potentials can lead to different band structures, even though the surface states are removed in both cases. This suggests that the passivation can introduce artifacts in the volume states, especially if the surface-to volume ratio of the structure is large. This indicates that a more careful passivation treatment is required.

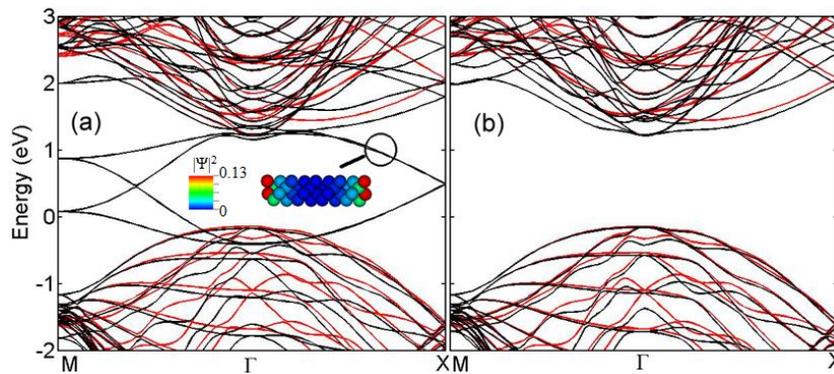


Fig. 1.2 (a) Band structures for a 2.2nm thick Si quantum well with (red) and without (black) passivation show that passivation removes the surface states in the band gap. The wave function of the surface state is shown in the inset. (b) Different passivation potentials lead to different band structures for the same quantum well suggesting that passivation can introduce artifact in the volume states.

1.2. Summary of This Work

In this work, a general lead method for contact self-energy calculation is developed. The method considers the realistic contact structures in the algorithm so that it can be applied to arbitrary contacts. This method is verified on a graphene nanoribbon device (GNR) with trumpet shape contacts and a $\text{Si}_{0.5}\text{Ge}_{0.5}$ nanowire transistor with alloy randomness in the contacts. A new passivation method based on the self-energy concept from NEGF is developed for arbitrary dangling bond scenario. The method is illustrated on Si quantum well structures with different SiO_2 oxidation configurations. The surface roughness scattering on black phosphorus transistors is investigated using these new methods.

This dissertation is organized as follows. Chapter 2 gives a general introduction to the NEGF and QTBM methods, and the existing contact self-energy methods. It also briefly introduces the empirical tight binding method (TB) and a new parameterization process. Chapter 3 introduces the algorithm improvement and implementation details of contact self-energy and quantum transport methods which give state-of-art performance. Chapter 4 presents the low rank approximation method (LRA) which aims at improving the efficiency of NEGF method and contact self-energy calculation by performing some approximations. Chapter 5 introduces the new general lead method and its validation and applications. Chapter 6 talks about the new passivation model and its applications. Chapter 7 introduces the new 2D material black phosphorus and the studies of transport properties in this material. Chapter 8 summarizes this work and provides an outlook of future research efforts.

2. QUANTUM TRANSPORT IN EMPIRICAL TIGHT BINDING

2.1. Introduction

Quantum transport approaches are gaining increasing interest in the device modeling community as the scaling of device dimension continues. A typical schematic of a device in transport modeling is shown in Fig. 2.1. The device channel is covered with oxide and the gate contact. The source and drain contacts are attached to the two ends of the device. The device channel is described with a Hamiltonian and a Schrödinger equation with open boundary condition is solved to obtain the charge density and current density. The open boundary conditions, which account for the charge injection and extract effects of the source and drain contacts, are usually calculated with contact self-energies. There are two typical methods to solve the open boundary Schrödinger equation: the NEGF [1]-[5] and the QTBM [6]-[9] methods. In the NEGF method, the Green's function of the device is solved by a matrix inversion and some matrix-matrix products. In QTBM, a linear equation is solved to obtain the wave function coefficients.

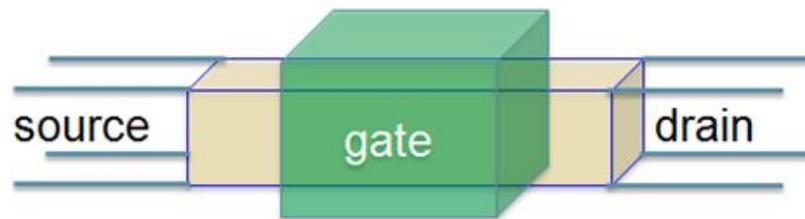


Fig. 2.1 A schematic of a typical device in transport modeling is shown. The device is gated in the central region and attached to two contacts at both ends.

The Hamiltonian, which is used to describe the interactions of atoms in the device, can be described with different methods, such as *ab-initio*, empirical tight-binding (TB), effective mass, etc. The TB method is widely used in nanodevice transport simulations [19]-[21] since it is accurate and efficient.

In this chapter, the quantum transport methods and the boundary condition methods are reviewed, introducing the basic concepts and equations. After that, the empirical TB method, the formulas and its typical notations as well as a new parameterization process are discussed.

2.2. Quantum Transport Methods

2.2.1. NEGF and RGF

The concept and the first applications of the NEGF method can be traced back to the early works of Schwinger [2], Kadanoff and Baym [3], Fujita [4], and Keldysh [5]. Datta [1] introduced the convenient formulations of NEGF in early 1990s and since then, the use of the NEGF method widely spread in device transport simulations. It allows the inclusion of complicated band structure models on atomistic resolution, and various kinds of scattering mechanisms under certain approximations [20]. Since only stationary NEGF formalism is used in this work, the discussions in the rest of this thesis are limited to it. More details of Green's function definitions in time domain and derivations of equations can be found in various literatures [1][20][22][23].

The equations of motion for the Green's functions including the effects of incoherent scattering are written as

$$(E - H_0 - \Sigma^R - \Sigma^{RB})G^R = 1 \quad (2.1)$$

and

$$(E - H_0 - \Sigma^R - \Sigma^{RB})G^< = (\Sigma^< + \Sigma^{<B})G^A \quad (2.2)$$

where $\Sigma^{R/<}$ represents various scattering mechanisms, such as phonon scattering, interface roughness scattering, impurity scattering, etc, and $\Sigma^{RB/<B}$ represents the self-energies due to the contacts. The Green's function $G^{R/<}$ is required to calculate the charge density and current density. The solution of (2.1) and (2.2) usually involves a self-consistent

calculation. In chapter 4.2 and chapter 7, the scattering due to acoustic phonon will be addressed. Here, only ballistic transport is discussed. Under the ballistic limit, $\Sigma^{R/<}$ will disappear from (2.1) and (2.2), so the equations can be simplified as

$$(E - H_0 - \Sigma^{RB})G^R = 1 \quad (2.3)$$

and

$$G^< = G^R \Sigma^{<B} G^A \quad (2.4)$$

To obtain G^R in (2.3) involves a matrix inversion, which is very expensive and heavy in memory. Consequently, usually a recursive method is applied to reduce the computational burden. In the recursive Green's function method (RGF) [20] only a few blocks of matrix G^R are solved so that it largely reduces the computational effort. The RGF equation can be written as

$$\mathbf{g}_{i,i}^r = (E - H_{i,i} - t_{i,i-1} \mathbf{g}_{i-1,i-1}^r t_{i-1,i})^{-1} \quad (2.5)$$

$$\begin{aligned} G_{N,N}^R &= (E - H_{N,N} - t_{N,N-1} \mathbf{g}_{N-1,N-1}^r t_{N-1,N} - \Sigma_D)^{-1} \\ G_{i,i}^R &= \mathbf{g}_{i,i}^r + \mathbf{g}_{i,i}^r t_{i,i+1} G_{i+1,i+1}^R t_{i+1,i} \mathbf{g}_{i,i}^r \\ G_{i,N}^R &= -\mathbf{g}_{i,i}^r t_{i,i+1} G_{i+1,N}^R \end{aligned} \quad (2.6)$$

where i denotes the index of matrix blocks, H and t represent the onsite block and coupling between two adjacent blocks of the device Hamiltonian matrix. The matrix \mathbf{g}^r represents the Green's function with contact at one side only. Solution of RGF involves two path iterations: once from source contact to the drain contact, which gives the \mathbf{g}^r matrix blocks; the other from drain contact to source contact, which gives the diagonal blocks and the right most column blocks of G^R . If only the transmission is needed, only G^R of the last slab N is sufficient.

The $G^<$ can be written as

$$-iG_{i,i}^< = f_d A_{i,i}^d + f_s (A_{i,i} - A_{i,i}^d) \quad (2.7)$$

$$\begin{aligned} A_{i,i} &= i(G_{i,i}^R - G_{i,i}^A) \\ A_{i,i}^d &= G_{i,N}^R \Gamma_{N,N}^d G_{N,i}^A \end{aligned} \quad (2.8)$$

where A and A^d are the spectral functions, which represent the density of states (DOS) of the whole device and the DOS due to the injection of drain contact, respectively. The charge density is the diagonal of $G^<$. Implementation details and optimizations made will be discussed in chapter 3.2.

2.2.2. QTBM

The QTBM methods can be traced back to the early works of Lent and Kirkner [7], and Ando [8]. Extensions of the QTBM equations to multi-band Hamiltonians were done by Khomyakov [9], and Luisier [6].

The QTBM method is applied to ballistic transport only, and can be proven to be physically equivalent to NEGF under the ballistic limit [6]. However, the QTBM method is found to be more efficient than NEGF/RGF for ballistic transport simulation, typically $\sim 5x$ faster in speed. The QTBM equation is a linear equation which can be written as

$$LHS \cdot \Psi = RHS \quad (2.9)$$

where Ψ is the solution, which represents the wave function coefficient of the device. All observables such as charge density and current density are calculated with the use of Ψ . The transmission is solved with

$$T = 2\pi i \cdot \text{trace}(\Psi_N^\dagger (\Sigma_D - \Sigma_D^\dagger) \Psi_N) \quad (2.10)$$

where N denotes the last matrix block, and Σ_D is the self-energy of the drain contact. The DOS is calculated with

$$DOS(i, j) = |\psi_{i,S;j}|^2 + |\psi_{i,D;j}|^2 \quad (2.11)$$

where i donates the index of slabs, j donates the electronic orbital, and S, D denote the source and drain contacts, respectively.

The left-hand-side (*LHS*) of (2.9) is the inverse of the Green's function G^R , which is the part in the parenthesis of (2.3). The contributions of the right-hand-side (*RHS*) matrix exist only at the first and last row blocks, which are related to the injection of source and drain contacts. The contribution of the source contact can be written as

$$RHS_1 = -T_{10}\Phi_p^+ + T_{10}g_1^R(D_{00}\Phi_p^+ + T_{0,-1}\Phi_p^+e^{ik_p\Delta}) \quad (2.12)$$

Here, RHS_l denotes the first row block of the RHS matrix. The T matrices denote the coupling Hamiltonians between different slabs and D_{00} denotes the Hamiltonian of the source contact. The surface Green's function of the source contact is represented in g_l^R . The propagating modes Φ_p^+ and the corresponding phase factors $exp(-ik_p\Delta)$ of the source contact are solved with the transfer matrix method [6]. The positive sign (+) means the direction from contact to device. For the drain contact, a similar equation is solved. In order to obtain the correct transmission and DOS using (2.10) and (2.11), normalization of the modes is required. The rule for normalization is written as

$$\Phi^\dagger\Phi = \begin{cases} 1, & \text{decaying mode} \\ 1/2\pi v, & \text{propagating mode} \end{cases} \quad (2.13)$$

$$v = 2Im(e^{ik\Delta}\Phi^\dagger T_{-1,0}\Phi) \quad (2.14)$$

where v is the group velocity for the corresponding propagating mode. The optimization of constructing RHS and its implementation, as well as the efficient solution of (2.9) will be discussed in chapter 3.3.

2.2.3. Boundary condition: contact self-energy

Existing methods for contact self-energy calculations include the direct iterative method, the Sancho Rubio method [10] and the transfer matrix method [6]. These methods are developed based on the assumption of periodic and semi-infinite contacts, where one unit cell is repeated along the transport direction and periodicity is exploited in the algorithm.

The simplest and most straightforward method is based on the direct iteration between the surface Green's function and the self-energy, which is written

$$\begin{aligned} g_i^R &= (E - H_i - \Sigma_i^R)^{-1} \\ \Sigma_i^R &= T_{i,i-1}g_{i-1}^R T_{i-1,i} \end{aligned} \quad (2.15)$$

Since (2.15) can only be solved in finite iterations, effectively it terminates the semi-infinite contact at a finite distance. This termination generates reflection, which is

negligible only if the starting block of the iteration is far away from the device/contact boundary, or effectively means the number of iterations has to be large enough. Typically this method requires $>10^3$ iterations. This limitation prevents the usage of (2.15) in real applications, even though it is straightforward and easy to implement.

The Sancho Rubio method [10] can be considered as a smart reformulation of (2.15). It also solves the surface Green's function iteratively, but it iterates over contact blocks in an exponential manner, thus it can converge within very few iterations. The equations are written as

$$\begin{aligned}
 \alpha_i &= \alpha_{i-1}(E - \varepsilon_{i-1})^{-1} \alpha_{i-1} \\
 \beta_i &= \beta_{i-1}(E - \varepsilon_{i-1})^{-1} \beta_{i-1} \\
 \varepsilon_i &= \varepsilon_{i-1} + \alpha_{i-1}(E - \varepsilon_{i-1})^{-1} \beta_{i-1} + \beta_{i-1}(E - \varepsilon_{i-1})^{-1} \alpha_{i-1} \\
 \varepsilon_i^s &= \varepsilon_{i-1}^s + \alpha_{i-1}(E - \varepsilon_{i-1})^{-1} \beta_{i-1}
 \end{aligned} \tag{2.16}$$

In (2.16) E is the energy. For the first iteration, ε_0 is the Hamiltonian of the contact, α_0 and β_0 are the coupling Hamiltonian connecting the contact to the device. The idea of (2.16) is illustrated in Fig. 2.2. Assuming the semi-infinite contact contains many repeated blocks. At the beginning, it starts from α_0 and β_0 which are of the adjacent blocks. At the second iteration, it jumps from block 0 to block 2. Following the same procedure, at the n 'th iteration, it jumps over 2^n blocks. This exponential scheme ensures the fast convergence of the algorithm.

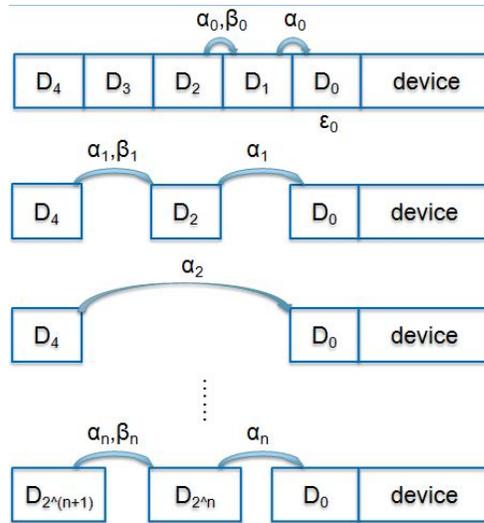


Fig. 2.2 The idea of iterative solutions in the Sancho Rubio method suggests an exponential extension of contact blocks which ensures fast convergence.

The iterations continue until α_i and β_i vanish, and then the surface Green's function is solved as

$$\mathbf{g}^R = (E - \varepsilon_i^s)^{-1} \quad (2.17)$$

The self-energy is then solved using the converged surface Green's function. The Sancho-Rubio method typically takes 30~40 iterations.

The other method is the transfer matrix method. This method is not an iterative method but is based on the solution of eigenmodes of the contact. With the periodic assumption of the contact, one can apply the Bloch theorem and assume some phase factors of the wave functions for different contact blocks, and then write down a matrix equation connecting different blocks. An improved version of the transfer matrix method [6] translates such a matrix equation into a normal eigenvalue problem as

$$M_2 \cdot \varphi_2 = \frac{1}{(e^{-ik\Delta} - 1)} \cdot \varphi_2 \quad (2.18)$$

where M_2 is a reformulation of the contact Hamiltonian, $\{\varphi_2\}$ are the eigenvectors. After solving equation (2.18), matrix-vector products are performed to obtain the full vectors.

$$\varphi_1 = (e^{-ik\Delta} - 1) \cdot M_1 \cdot \varphi_2 \quad (2.19)$$

The contact eigenmodes are constructed from the vectors φ_1 and φ_2 as $\Phi = \{\varphi_1, \varphi_2\}^\dagger$. The self-energy is calculated with the help of the contact modes.

$$\begin{cases} \tilde{\mathbf{g}}^R = (\Phi^\dagger D_{00} \Phi + \Phi^\dagger T_{0-1} \Phi e^{-ik\Delta})^{-1} \\ \Sigma^R = T_{10} \mathbf{g}^R T_{01} = T_{10} \Phi \tilde{\mathbf{g}}^R \Phi^\dagger T_{01} \end{cases} \quad (2.20)$$

In (2.20) D_{00} is the Hamiltonian of the contact unit cell, T_{01} is the coupling Hamiltonian, and $e^{-ik\Delta}$ are the phase factors of the contact modes. The implementation of (2.18)-(2.20) is not totally trivial, the performance for some specific cases can be improved significantly by making certain optimizations. These optimization schemes will be discussed in chapter 3.4. Note that to obtain the correct QTBM results one has to add a random potential to the contact Hamiltonian to break the degeneracy of the modes.

The above methods serve as the standard approaches for contact self-energy calculations in NEGF, and are widely used to study different types of nanodevices. However, they are based on the periodic and semi-infinite contact assumption so that they always inject charges with DOS from periodic contacts. As a result, they are not applicable for complicated contact geometries or contacts with randomness (alloy, roughness, etc).

Another method developed in recent years introduced a complex absorbing potential (CAP) in the contact Hamiltonian [24]. The self-energy is then calculated either by solving an eigenvalue problem or an iterative method. This method is proved to be very efficient and been widely applied to time dependent density functional based NEGF simulation [25]-[27]. Yet all these applications are still based on ideal periodic contacts.

2.3. Empirical Tight Binding Method

2.3.1. Motivation

Modern semiconductor devices consist of complicated geometries and composition of multiple materials. In addition to traditional materials like Si, Ge or GaAs, new materials such as transition metal dichalcogenides (TMD) [28] and black phosphorus [29] have joined the device community in recent years. Confined structures such as FINFETs modeled as ultra-thin bodies (UTB) or nanowires are often adopted in device design. An accurate description of the band structure for these structures is the prerequisite for any device simulation. The effective mass approximation works well in the vicinity of conduction band minima and valence band maxima, but doesn't always ensure a correct calculation of the quantization levels in nanostructures. The *ab-initio* methods which have no adjustable parameters can offer atomistic representations with subatomic resolutions for various materials and structures. However, it is generally too expensive for realistic devices which usually consist of $>10^4$ atoms. Also, it is an equilibrium theory which in principle cannot be applied in out-of-equilibrium conditions under which electronic devices are operating. During the last two decades the empirical TB methods have become the main choice of many researchers since these methods can reliably produce the band structures for nanostructures efficiently.

2.3.2. Tight binding Hamiltonian: formula and notation

The empirical TB method was firstly developed by Slater and Koster in the early 1950s [30]. Since then many research efforts have been invested in pushing forward the TB methods and applications [31]-[34]. Discussion of the TB method details is beyond the scope of this work. Here, only the fundamental concept, the basic formula to create the Hamiltonian, and the widely used notations are introduced.

For a given crystal structure, the TB wave functions can be written as a linear combination of localized atomic orbitals (LCAO) $\phi_n(r-R_i)$ with coefficients C_n . These functions are located on the various atoms i at position R_i , where n refers to the quantum number. With periodicity exists in the structure, the Bloch sum is written as

$$\psi(r) = \sum_{\alpha, \sigma, k} C_{\sigma}^{\alpha}(k) \sum_{R_i^{\alpha}} \phi_{\sigma}^{\alpha}(r - R_i^{\alpha}) e^{ikR_i^{\alpha}} \quad (2.21)$$

where α denotes the atom, σ denotes the orbital type, such as s , p , d , s^* , etc, and k is the wave vector. The matrix elements of the TB Hamiltonian can be obtained by applying the two center approximation and orthogonalizing the orbital functions. The equation for TB in matrix form can be written as

$$(H^{\alpha} - E(k))C^{\alpha}(k) + \sum_i V^{\alpha\beta}(b_i^{\alpha})C^{\beta}(k)e^{ikb_i^{\alpha}} = 0 \quad (2.22)$$

In (2.22) H^{α} and $V^{\alpha\beta}$ are matrices considering all orbitals σ . The dimensions of these matrices depend on the degrees of freedom of the TB model. For example, for $sp^3d^5s^*$ model the dimension is 10 without spin, or 20 with spin. H^{α} is the onsite term of an atom α . It is a diagonal matrix where the elements denote the onsite energy of s , p , d , s^* orbitals [30]. The off-diagonal elements are 0 except when spin is considered. The interatomic interactions between atom α and atom β are represented in $V^{\alpha\beta}$. The sum in (2.22) loops over all neighbors of atom α . The matrix element in $V^{\alpha\beta}$, for example, the overlap between orbital p_x^{α} and d_{xy}^{β} is approximated in terms of two integrals, the one for σ and the one for π . The matrix elements can be written as [30]

$$V_{x,xy}^{\alpha\beta}(b_i^\alpha) = \sqrt{3}l^2 m V_{pd\sigma} + m(1-2l^2)V_{pd\pi} \quad (2.23)$$

The direction cosines l , m , n describe the bond pointing from atom α to atom β . All the matrix elements of $V^{\alpha\beta}$ can be found in [30]. For UTB or nanowire where periodicity is broken along certain directions, the construction of the Hamiltonian is similar, except that at the surface a hard-wall boundary is applied and the atoms need to be passivated. The passivation of the surface atoms are discussed in chapter 6.

The notation for TB parameters of [34] is adopted in this work. For $sp^3d^5s^*$ TB model, the notation is listed in Table 2.1.

Table 2.1 Notation of TB parameters

Notation	Physical term	Notation	Physical term
E_s	onsite of s orbital	$V_{ss\sigma}$	coupling $ss\sigma$
E_p	onsite of p orbital	$V_{sp\sigma}$	coupling $sp\sigma$
E_{s^*}	onsite of s^* orbital	$V_{sd\sigma}$	coupling $sd\sigma$
E_d	onsite of d orbital	$V_{ss^*\sigma}$	coupling $ss^*\sigma$
$V_{ps^*\sigma}$	coupling $ps^*\sigma$	$V_{s^*s^*\sigma}$	coupling $s^*s^*\sigma$
$V_{pp\sigma}$	coupling $pp\sigma$	$V_{s^*d\sigma}$	coupling $s^*d\sigma$
$V_{pp\pi}$	coupling $pp\pi$	$V_{dd\sigma}$	coupling $dd\sigma$
$V_{pd\sigma}$	coupling $pd\sigma$	$V_{dd\pi}$	coupling $dd\pi$
$V_{pd\pi}$	coupling $pd\pi$	$V_{dd\delta}$	coupling $dd\delta$

2.3.3. Brief introduction to the TB mapping method

The accuracy of the TB methods depends on the careful calibration of the parameters as described above. The traditional way to determine the TB parameters is to fit the TB results to the experimental data of bulk materials [19][31]. However, this traditional fitting process has a potential problem; the parameters obtained from this process might introduce ambiguity when applied to confined structures like UTBs or nanowires. It has been shown in [35] the application of existing TB parameters to As terminated GaAs

UTB leads to unphysical valence states. Another problem is the missing of explicit basis functions in this fitting process. This makes it hard to predict wave function dependent quantities like the optical matrix elements.

To tackle these problems, a new parameterization method is developed by Tan *et al.* in [36]. In this method, the TB parameters are optimized by mapping the TB results, such as dispersions and wave functions, to *ab-initio* results. The validation of wave functions in this method provides additional targets for TB parameterization which reduces the arbitrariness of the parameters. As a result, the TB parameters are more transferable and reliable. This mapping process is shown in Fig. 2.3. It is an iterative process which fits the TB dispersions and wave functions to *ab-initio* calculations by optimizing the parameters. Once a certain precision is matched, the process finishes and the TB parameters are obtained. In this work, this mapping process is applied to the parameterization of the passivation model as described in chapter 6, as well as the TB parameters of black phosphorus in chapter 7.

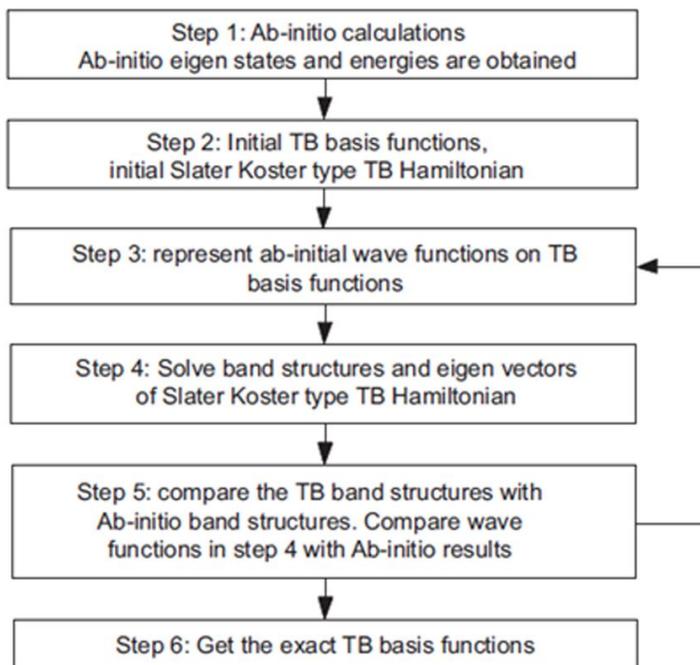


Fig. 2.3 The process of mapping *ab-initio* results to TB

2.4. Summary

In summary, this chapter introduces the basic concept and formulas for the NEGF and QTBM methods. The widely used contact self-energy methods in literatures are discussed. The empirical TB method, including the construction of the Hamiltonian, the notation of parameters, as well as the parameterization method is discussed.

3. ALGORITHM IMPROVEMENT AND IMPLEMENTATION

3.1. Introduction

As the dimension of electronic devices is shrinking and approaching ballistic limit, quantum effects such as tunneling, confinement and interference become crucial in device performance. Classical transport approach based on Boltzmann Transport Equation (BTE) cannot represent these quantum effects accurately. Consequently, quantum transport models gain increasing importance in device modeling and simulation. Algorithms such as NEGF [1] and QTBM [6] provide a general framework for quantum transport and are therefore accepted for modeling the physics of nanoscale devices [6][12]-[14][21]. However, these algorithms involve numerically expensive matrix operations such as eigenvalue problems, matrix inversions and matrix-matrix products. For a Si nanowire FET with 3nm diameter, 20 nm length in 10 band TB model, the device contains ~20 000 atoms. Solving an I - V characteristic with 10 bias points for such a device requires ~100 000s. Consequently, for realistic device simulations efficient implementations of these algorithms are critical. Although the NEGF and the QTBM algorithms are thoroughly discussed in literatures [1]-[9], details of efficient implementations of these algorithms are rarely given. In this chapter, the details of these algorithms are discussed and their efficient implementation into the multi-scale simulation tool NEMO5 [37] is presented. The impacts of efficient implementations are illustrated on a 20nm long, 3nm thick Si nanowire in 10 band atomistic TB representation. Performance improvements of QTBM and NEGF for time and peak memory of factors of 3-5 over the current state of literatures can be achieved with the presented implementation details.

3.2. RGF Implementation

The NEGF method requires the solution of the retarded Green's function (G^R) and lesser Green's function ($G^<$) in the device to obtain the transmission and the charge density. The key operation of the NEGF method is the inversion of a matrix with the same rank as the device Hamiltonian. The solution time and the peak memory usage increases dramatically as the device dimension increases. The recursive Green's function method (RGF) [20] is well-known for improving the efficiency of NEGF calculation. It allows solving the transmission and the charge density with only a minimum number of blocks of the G^R matrix.

As shown in chapter 2.2.1 the RGF algorithm divides the device into slabs along transport direction and solves the relevant G^R blocks recursively. Afterwards $G^<$ matrix is solved to obtain the charge density. It requires to store the following matrices: 1) the diagonal blocks of the retarded Green's function g^r for forward iterations, 2) the diagonal blocks and the last column blocks of the retarded Green's function G^R for backward iterations, and 3) the diagonal blocks of the lesser Green's function $G^<$. These matrix blocks are dense, thus for realistic structures the memory consumption to store these blocks are huge. Fortunately, it is not necessary to store all these three matrices. As shown in (2.5), it always requires to store the g^r blocks since the backward path in (2.6) relies on it. Fig. 3.1(a) shows the forward iterations. However, as shown in (2.6) during the backward path, the i 'th diagonal block and column block of G^R only depend on the i 'th block of g^r . After G^R is solved for slab i , the corresponding g^r block is not needed and can be de-allocated. Furthermore, (2.7) and (2.8) show that the A matrix and $G^<$ matrix can be solved for each slab i during the backward iterations of RGF. After $G^<$ is solved for slab i , the corresponding G^R blocks (both diagonal and off-diagonal) are not needed anymore and can be de-allocated. Also, since only the diagonal of $G^<$ is required for the charge density, the storage of the whole diagonal block of $G^<$ can be avoided. Consequently, during the backward iterations no extra matrix blocks except for the diagonal elements of $G^<$ are stored, so that the peak memory of RGF algorithm is dominated only by g^r blocks in the forward iterations. Fig. 3.1(b) shows this process.

Table 3.1 shows that with these improvements the peak memory is minimal and does not increase significantly with the number of energy points.

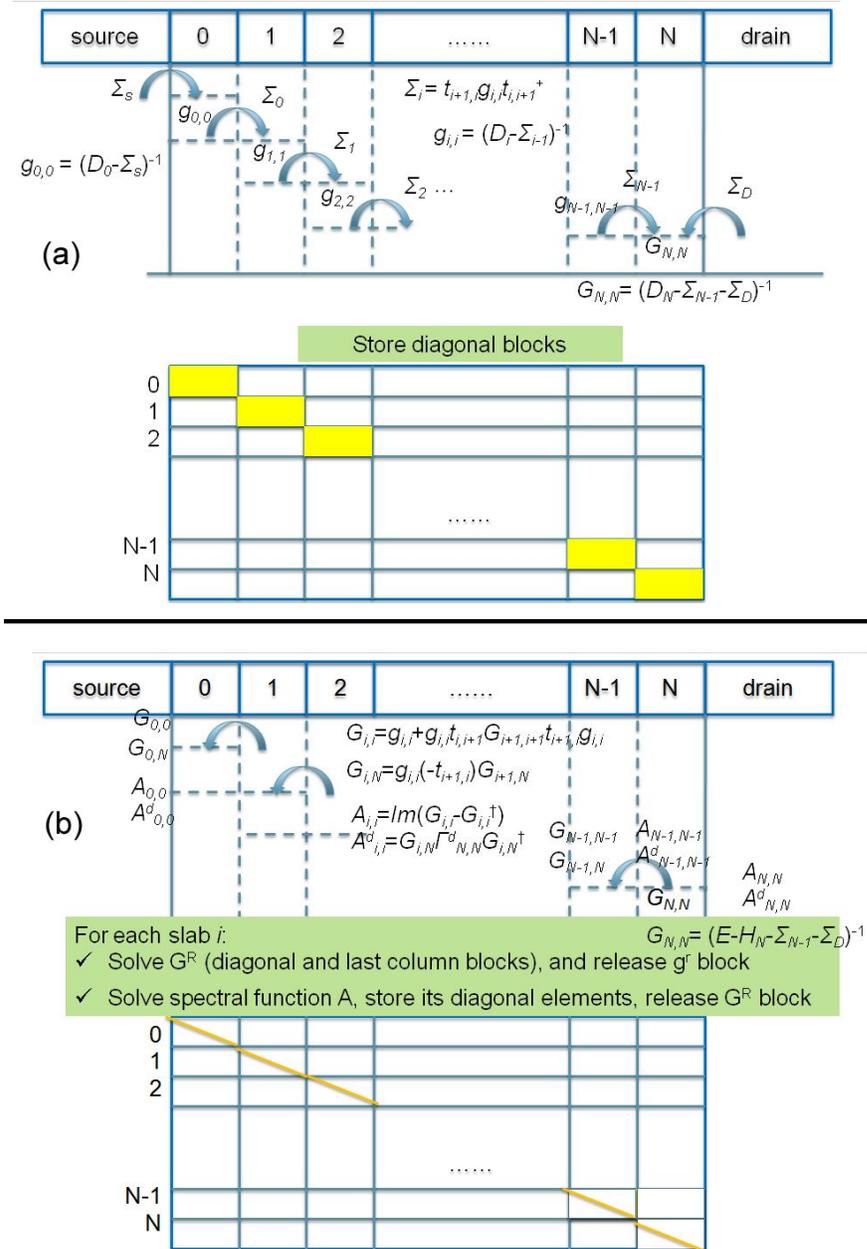


Fig. 3.1 RGF process for (a) forward iterations and (b) backward iterations. Yellow blocks denote the matrix elements that are stored. The yellow lines in (b) indicate that only the diagonal elements are stored during the backward iterations.

Table 3.1 Peak memory usage for RGF in Gigabytes. *std*, the state of literature, *opt*, the optimized way in this work. It shows that the peak memory does not increase significantly with more energy points with the stated optimizations.

	std	opt
1 energy point	4.56	1.63
3 energy points	13.8	1.78
5 energy points	22.95	1.89

3.3. QTBM Implementation

The QTBM method requires the solution of a linear equation to obtain the propagating wave functions in the open device. As mentioned in chapter 2.2.2, the *LHS* of this linear equation is just the device Hamiltonian attached with the contact self-energies from the two contacts. The *RHS* of the equation represents the charge injection from the contacts, which is usually described by the contact propagating modes Φ_p , the phase factors $e^{ik\Delta}$ and the surface green's function g^R . The solution of the QTBM equation represents the propagating wave functions of the device. These wave functions are used to solve the transmission and the charge density. The hotspots of the QTBM method are: 1) the construction of the right hand side matrix of the QTBM equation and 2) the solution of the linear QTBM equation.

The *RHS* equation in (2.12) requires the explicit solution of g^R . One can avoid that by inserting (2.20) into (2.12) and using the fact that the contact domain is periodic:

$$T_{0,1}\Phi^-e^{-ik\Delta} + D_{00}\Phi^- + T_{0,-1}\Phi^-e^{ik\Delta} = 0 \quad (3.1)$$

where the minus sign (-) means the direction from device to contact. The *RHS* can then be simplified:

$$RHS_1 = -T_{10}\Phi_p^+ - \Sigma^R\Phi_p^+e^{-ik_p\Delta} \quad (3.2)$$

where Σ^R is the self-energy in (2.20). It can be shown that (3.2) has much fewer matrix products compared to (2.12), which is easier to implement and more efficient. Another important point is that (3.2) avoids solving g^R explicitly, which allows some improvement when solving (2.20). This improvement will be discussed shortly in chapter

3.4. A speed up of $\sim 35x$ for the construction of the *RHS* is observed by using (3.2) instead of (2.12).

The solution of the linear equation can be efficient by considering the sparsity pattern of the *LHS* matrix. Since the *LHS* of the QTBM equation is the device Hamiltonian plus the self-energies of the two contacts, it is a very sparse matrix except for two small dense blocks at the upper left and lower right matrix corner. MUMPS [38] is found to be very efficient for factorizing this matrix, thus it is often used as the preconditioner for the linear equation. The device can be partitioned into several slabs along the transport direction so that the *LHS* matrix is divided into several matrix blocks. In this way, the linear QTBM equation can be solved spatially (block) distributed in parallel. Furthermore, for nanowires without explicit spin-orbit coupling, the elements in the center blocks of the *LHS* matrix are real, so that these blocks can be solved with real-type operations. Fig. 3.2 shows the scaling performance of QTBM calculations for 20nm Si nanowire with different cross-sections using NEMO5. It suggests that the scaling is close to ideal scaling performance up to 4 cores, while above that the scaling becomes weaker due to the increasing MPI communications.

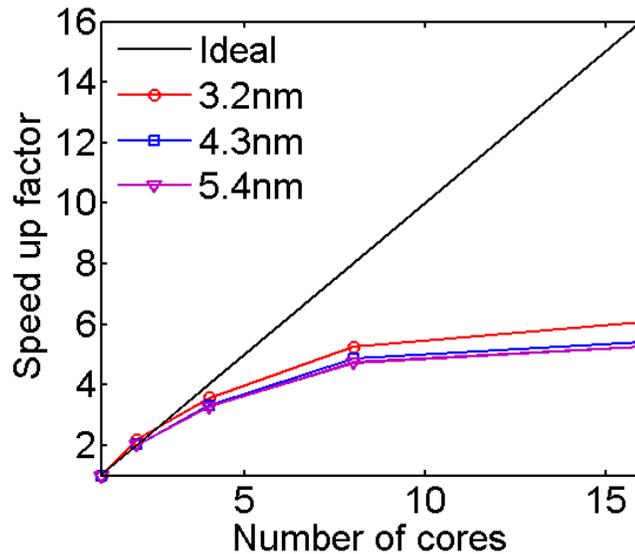


Fig. 3.2 Speed up factor of NEMO5 QTBM calculations with respect to number of cores used, for 20nm long Si nanowires with different diameters. It suggests that the calculations are close to ideal scaling performance up to 4 cores.

3.4. Implementation of Transfer Matrix Method

In the QTBM or NEGF simulations a critical step is to solve for the open boundary condition, which is represented by the contact self-energies. As discussed in chapter 2.2.3, there are several known self-energy algorithms, such as the Sancho-Rubio method [10] and the transfer matrix method [6]. The Sancho-Rubio method is based on an iterative solution of the surface Green's function. The transfer matrix method is based on a generalized eigenvalue problem for contact modes and translation of the modes into a surface Green's function and a contact self-energy. A modified version of the transfer matrix method presented in [6] transforms the generalized eigenvalue problem into a normal eigenvalue problem to reduce the numerical load of the contact self-energy calculations. Both methods are implemented in NEMO5, but only the more efficient transfer matrix method is discussed here. There are four numerical hotspots of this algorithm: 1) translation of the generalized eigenvalue problem into a normal eigenvalue problem; 2) solution of the normal eigenvalue problem; 3) matrix-vector products to obtain the contact modes; 4) matrix-matrix products to obtain the contact self-energy.

The transformation of the generalized eigenvalue problem into a normal eigenvalue problem is written as

$$M = (H - P)^{-1}P \quad (3.3)$$

where H and P are just some simple modifications of the contact Hamiltonian as shown in [6], and M is the matrix whose eigenvalue problem will be solved. A straightforward implementation of (3.3) involves an inversion and a matrix-matrix product of complex matrix type. In NEMO5, (3.3) is rewritten as a linear equation

$$(H - P) \cdot M = P \quad (3.4)$$

The M matrix of (3.4) can be obtained by solving a linear equation instead. It is important to mention that for electrons in nanowire structures without explicit spin-orbit coupling in the TB representation, all the matrix elements of the Hamiltonian are real. As a result, all matrices in (3.4) can be solved with real type matrix operations rather than

complex type. Table 3.2 shows a speed up of $\sim 6x$, by solving (3.4) in real type operations instead of solving (3.3) in complex type.

As shown in (14) and (15) of [6], the transformation of the generalized eigenvalue problem into a normal eigenvalue problem results in the reduction of the actual matrix equation size. The relevant eigenvalue problem to be solved is written in (2.18), where M_2 is the lower right block of M matrix in (3.3). Similar to the above step, the M_2 matrix is a real matrix which allows the usage of a real type eigensolver (Lapack degeev in this work) [39]. Table 3.2 shows that this gives a speed up of about $4.3x$ compared to calling the complex type eigensolver.

After the solution of the eigenvalue problem, the contact modes are calculated with (2.19). In (2.19), the M_1 matrix is the upper right block of the M matrix, which is again a real matrix. Consequently (2.19) is a real matrix-complex vector product. However, the eigenvectors $\{\varphi_2\}$ in (2.18) from the real type Lapack eigensolver are combinations of some real vectors $\{\psi\}$ with the following rules [39]:

If the j 'th eigenvalue is real, it holds

$$\varphi_2(j) = \psi(j) \quad (3.5)$$

If the j 'th and the $(j+1)$ 'th eigenvalues form a complex conjugate pair, it holds

$$\varphi_2(j) = \psi(j) + i \cdot \psi(j+1) \quad (3.6)$$

This allows first performing the product between the real matrix M_1 and the real vector $\{\psi\}$, and then combining the result vectors to generate $\{\varphi_1\}$ following the rules described above. This leads to a speed up of $\sim 12x$ compared to a direct solution of (2.19) as shown in Table 3.2.

The solution of the contact self-energy requires the surface Green's function g^R and the contact modes $\Phi = \{\varphi_1 \varphi_2\}^\dagger$ as shown in (2.20). A couple of matrix-matrix products and a matrix inversion are involved in (2.20). However, if only Σ^R is required, the explicit solution of g^R can be avoided such that (2.20) can be rewritten as a linear equation

$$\begin{cases} (\Phi^\dagger D_{00} \Phi + \Phi^\dagger T_{0-1} \Phi e^{-ik^{-}\Delta}) \cdot X = \Phi^\dagger T_{01} \\ \Sigma^R = T_{10} \Phi \cdot X \end{cases} \quad (3.7)$$

Since T_{0l} is very sparse and Φ is usually a rectangular matrix, solving the linear equation in (3.7) is much more efficient than the matrix inversion and products in (2.20). A speed up of about 5x is achieved compared to an explicit solution of g^R in (2.20).

The total speed up of $\sim 5x$ of the transfer matrix method with all the optimizations described above is shown in Table 3.2.

Table 3.2 Timing comparison in seconds, for 1 energy point. std, the state of literature, opt, the optimized way discussed in this work. It shows that the stated optimizations lead to a total speed up of 5x in the contact self-energy calculations and 35x in the construction of the QTBM RHS matrix.

Chapter	std	opt	std/opt
3.4 1)	11.7	2	5.9
3.4 2)	63.4	14.8	4.3
3.4 3)	12.5	1	12.5
3.4 4)	12	2.5	4.8
Σ total	99.6	20.3	4.9
3.3 1)	24.3	0.7	34.7

3.5. Summary

In summary, the optimized implementations of NEGF/RGF, QTBM and contact self-energy methods in NEMO5 are discussed. The minimal consumption of memory in RGF is achieved by controlling the storage of the matrix blocks during the backward iterations. In the QTBM and the transfer matrix methods, several optimization steps such as replacing complex type matrix operations with real type, avoiding the explicit solution of surface Green's function and reformulation of the equations are shown, and a speed up of 5x is achieved for the nanowire calculations.

4. LOW RANK APPROXIMATION METHOD

4.1. Introduction

As discussed in the above sections, there are two widely used quantum transport methods: the NEGF method and the QTBM method. The QTBM method is efficient for the ballistic transport problem, while the described framework does not allow the treatment of scattering. The NEGF method, however, is applicable to transport problem with various scattering mechanisms, such as electron-phonon scattering, impurity scattering, etc. As a result, the NEGF method is with more interest in device simulation since today's realistic devices have not reached the ballistic limit.

As mentioned in chapter 2 and 3, the numerical burden for solving NEGF equations is huge. There are several approximations to reduce the numerical costs for particular devices and situations. The RGF method [20] reduces the peak numerical burden to a device dependent sub-block matrix of the system's Hamiltonian, and the computational cost scales linearly with the number of blocks but cubically with the block size. It is widely used for the simulations with one transport direction such as FinFETs and nanowire structures [21]. Mode space methods [40] in similar wire structures as well as the newly developed equivalent transport mode method [41] separate the transport direction from transverse confinement directions to reduce the block size in each layer. These methods require a clear distinction between the transport direction and the transverse directions. When this distinction becomes blurred, as in the case of incoherent scattering, the numerical efficiency drops significantly. Another way to solve ballistic NEGF equations efficiently is the contact block reduction method (CBR) [42][43]. However, this method cannot handle the incoherent scattering problem and has been shown to have limited capabilities in TB basis sets [44].

In this chapter, the low rank approximation method (LRA) is applied to the NEGF equations in the presence of inelastic electron-phonon scattering. This method is an extension of the basis reduction method proposed by Greck [45]. LRA is a general mathematical concept and the key idea of it is to keep the critical information of a matrix and drop the less important ones. In the framework of NEGF, the transport problem is transformed from the original basis into a reduced basis. The matrix rank of the reduced basis is much smaller than the original one so that the numerical load in solving the equations in the reduced basis is much lower.

In this chapter, the LRA method in the effective mass approximation and the empirical TB representation will be discussed separately. The LRA in the contact self-energy calculation will be introduced also.

4.2. LRA in Effective Mass Approximation

4.2.1. Reduced real space basis

The NEGF calculations are time consuming since they involve the inversion and matrix product of matrices with the rank N of the system's Hamiltonian. The fundamental idea of LRA is to reduce the numerical cost by transforming the NEGF equations into a space of lower rank n and solving the equations therein. It is expected that the closer the basis functions of the reduced space are to the physically relevant quasiparticles of the device, the better the LRA approximation is and the smaller the ratio n/N can be used.

The method is demonstrated on the stationary vertical transport in laterally homogeneous quantum well heterostructures that are in contact with two charge reservoirs. The electron structure is represented in terms of a single band effective mass Hamiltonian H_0 that is represented in a basis of N position eigenfunctions

$$H_0 = \frac{-\hbar^2}{2} \frac{d}{dz} \frac{1}{m^*(z)} \frac{d}{dz} + \frac{\hbar^2 k_{\parallel}^2}{2m^*(z)} + V(z) \quad (4.1)$$

where k_{\parallel} is the in-plane momentum and $V(z)$ denotes a position dependent potential. In stationary NEGF the equations are [23]:

$$\begin{aligned}
G^R &= (E - H_0 - \Sigma^R)^{-1} \\
G^< &= G^R \Sigma^< G^{R\dagger} \\
\Sigma^< &= G^< D^< \\
\Sigma^R &= G^R D^R + G^R D^< + G^< D^R
\end{aligned} \tag{4.2}$$

Here, the retarded and lesser Green's functions are given by G^R and $G^<$. D is the sum of all environmental Green's functions that incorporate, e.g. phonons, and Σ denotes the self-energies. In this chapter, the inelastic scattering is dominated by the longitudinal acoustic phonons given by scattering self-energies [20][23]

$$\begin{aligned}
\Sigma_{ac}^{<,R}(z, z', k_{\parallel}, E) &= \frac{1}{(2\pi)^3} \frac{k_B T D_{ac}^2}{2\rho v_s^2} \\
&\times \int d\vec{q}_{\parallel} dq_z e^{iq_z |z-z'|} \tilde{G}^{<,R}(z, z', |\vec{k}_{\parallel} - \vec{q}_{\parallel}|, E)
\end{aligned} \tag{4.3}$$

with the energy-average Green's functions

$$\tilde{G}(z, z', q_{\parallel}, E) = \frac{1}{2\hbar\omega_q D_{ac}} \int_{E-\hbar\omega_d}^{E+\hbar\omega_d} dE' G(z, z', q_{\parallel}, E') \tag{4.4}$$

The acoustic deformation potential and the material density are denoted by D_{ac} and ρ , and the acoustic phonon frequency, sound velocity, and Debye frequency are denoted by ω_q , v_s and ω_d .

The first step of the LRA method is to find the basis functions. The n eigenfunctions of the free particle Hamiltonian H_0 with Neumann boundary conditions are solved with the smallest eigenenergies E_i

$$H_0 \phi_i = E_i \phi_i, \quad i = 1, 2, \dots, n \tag{4.5}$$

Here, n is chosen such that the energy E_n is about several $k_B T$ above the highest chemical potential of all contacts. Consequently, all quasi-particles with energies below E_n are appropriately considered in the calculation. This is essential to capture all occupied states in order to predict the density accurately. It is worth to mention that (4.5) is sufficient if only ballistic transmission and density is required. In order to solve inelastic scattering, the following steps are required. The n orthonormal eigenstates ϕ_i are set into the n

columns of a $N \times n$ dimensional matrix S . To define the locality/non-locality of scattering self-energies, the position operator of the real space discretization X is transformed into the reduced rank basis

$$X_S = S^\dagger X S \quad (4.6)$$

The position operator X is a diagonal matrix, whereas the reduced rank matrix X_S is a dense matrix. Therefore, the operator X_S is diagonalized to find the reduced rank position eigenfunction basis $\{\psi_i\}$

$$X_S \psi_i = x_i \psi_i, \quad i = 1, 2, \dots, n \quad (4.7)$$

These orthonormal basis functions $\{\psi_i\}$ define the columns of a squared, unitary $n \times n$ transformation matrix P . The NEGF equations in the reduced basis $\{\psi_i\}$ can be written as

$$\begin{aligned} G_P^R &= (P^\dagger S^\dagger (E - H_0) S P - \Sigma_P^R)^{-1} \\ G_P^< &= G_P^R \Sigma_P^< G_P^{R\dagger} \\ \Sigma_P^< &= G_P^< D_P^< \\ \Sigma_P^R &= G_P^R D_P^R + G_P^R D_P^< + G_P^< D_P^R \end{aligned} \quad (4.8)$$

The basis function ψ_i is associated with a position x_i such that the equations above are discretized in a reduced rank real space representation. Position dependent scattering self-energies such as the acoustic phonon scattering self-energy in (4.3) are solved self-consistently in the reduced real space. The introduction of the reduced real space avoids the back-transformation of the Green's function into the original real space in the self-consistent calculation when position dependent scattering self-energy is calculated.

Once the calculations of (4.8) are converged, the diagonal and first off-diagonal elements of $G^<$ are transformed back into the original real space to solve for the charge density and current density. The accuracy of the Green's function relies on the rank of the reduced space n . It will be shown later that the smaller n is compared to N , the more unreliable the LRA results are. The LRA approximated current density usually shows an oscillation in the original space. To predict current voltage characteristics in the LRA method, this inhomogeneous current density is averaged over the device excluding areas

within $\Delta N/n$ from the contacts, where Δ denotes the average mesh point distance in the original real space representation.

4.2.2. Comparison with existing methods

It is important to highlight some differences of this method with other well established efficient NEGF algorithms such as the CBR method [42][43], mode space approaches [40][41] and the RGF method [20].

In the CBR method, the NEGF equations are first transformed into an efficient representation to utilize the fact that ballistic calculations require only some sections of the retarded Green's function G^R to be solved. A rectangular transformation that reduces the rank of the NEGF equations is applied only after that first transformation. Although the CBR method is very efficient, it is fundamentally limited to ballistic calculations and has limited capabilities in TB basis [44].

The mode space approach [40] assumes a separation Ansatz for the wave functions of propagating quasi-particles. Typically, the Ansatz requires confined modes or plane waves perpendicular to the transport direction. The mode space approach allows a significant rank reduction of the NEGF equations. The computational burden is even further reduced if these modes are well separated in energy and the particle's propagation does not couple different modes. However, if the device contains inhomogeneities, such as impurities, non-conformal confinement, etc, the number of the modes is no longer a conserved quantum number. Then the modes are coupled and the rank of the mode space has to be large to predict transport without loss of accuracy.

The RGF method allows limiting the calculation of the retarded Green's function to selected parameter intervals of the propagation space. This allows limiting ballistic NEGF calculations on the required elements of G^R only, which results in much faster transport solutions than the case when the complete G^R is solved. NEGF that includes incoherent nonlocal scattering, however, requires the full G^R which deteriorates the advantages of this recursive method. In contrast to these three methods, the LRA method allows the inclusion of any incoherent scattering as well as arbitrary device geometries. The basis functions of the LRA method are device dependent wavefunctions that

automatically include non-conformal confinement, if such confinement appears. A timing comparison with all these efficient methods is shown in Table 4.1. The calculations are performed on a 100nm long GaAs resistor in the effective mass approximation. All the efficient algorithms show speed ups compared to the original NEGF equation. The LRA method shows a similar performance compared to the CBR method in this example.

Table 4.1 The measured time consumptions of ballistic transport calculations for a 100nm long GaAs resistor with different methods. The results are solved in MATLAB in serial. Speed ups are observed in all the established efficient NEGF algorithms. The LRA method shows a similar performance compared to the CBR method in this example.

Direct Inversion	RGF	CBR	Mode Space	LRA
99.6s/1x	42s/2.4x	19s/5.2x	36/2.8x	20s/5x

4.2.3. Numerical complexity and memory usage analysis

Coherent quantum transport calculations for realistically extended devices have been shown to efficiently consume the computational power of over 220 000 processing cores [46]. Incoherent NEGF based calculations require about 100x more computational power, which are limited to generally unrealistic small structures, and can only scale to ~100 000 cores [47]. Involving incoherent scattering in realistically extended devices requires dramatically large computational resources.

The numerical complexity and memory usage of self-consistent NEGF calculations reduce when the LRA method is applied. To quantify that, this section compares the number of floating point operations and the memory usage of the NEGF calculation with an approximated ballistic NEGF solution that employs the LRA method. In the following, the transport problem is assumed to be originally discretized with N orthogonal basis functions. Within the LRA method, the rank of the NEGF equations is reduced down to n . The energy and other conserved quantum numbers of the NEGF equations are resolved with a mesh of N_E points. To get a conserved current density within the self-consistent Born approximation, N_i iterations of the Green's functions and self-energies are required. The integral in (4.3) is solved with N_{ph} energy points.

The exact solution of one retarded Green's function involves the inversion of an N dimensional matrix which requires $O(N^3)$ floating point operations. The solution of each lesser Green's function involves two matrix-matrix products with a numerical load of $O(N^3)$ floating point operations. The solution of the local scattering self-energy of (4.3) is $O(N_{ph} \times N)$ for each energy point in each iteration. In total, solving the NEGF equations exactly requires $N_E \times N_i \times [O(N^3) + O(N_{ph} \times N)]$ floating point operations, while the memory needed to store the matrix representation of Green's functions and self-energies is $N_E \times O(N^2)$ floating point numbers.

The LRA method can be decomposed into three steps: 1) the transformation of the NEGF equations into the reduced space, 2) the solution of the NEGF equations within the reduced space and 3) the back-transformation of some relevant results into the original space. Step 1) requires first to get the eigenstates that construct the transformation matrix, i.e. $O(N^2 \times n)$ floating point operations. The memory used to store the transformation matrix is $O(N \times n)$ floating point numbers. The transformation of the device's Hamiltonian into the reduced space requires then $O(N^2 \times n)$ floating point operations. The contact self-energies have to be transformed for every energy point at each iteration. Since the contact self-energy is zero except for the mesh points adjacent to the contacts in the original space, each transformation requires only $N_E \times N_i \times O(n^2)$ floating point operations and $N_E \times O(n^2)$ floating point numbers to be stored. Solving the NEGF equations in the reduced space in step 2) requires $N_E \times N_i \times O(n^3)$ floating point operations and memory usage of $N_E \times O(n^2)$ floating point numbers. The calculation of the acoustic phonon self-energy costs $O(N_{ph} \times n)$ operations in the reduced real space for each energy point at each iteration. To calculate the energy resolved densities and current densities in the original space, the step 3) requires to back-transform the diagonal and the first off-diagonals of $G^<$ of the original space. This transformation requires $N_E \times O(N \times n^2)$ floating point operations and memory usage of $N_E \times O(N)$ floating point numbers.

In a typical effective mass NEGF calculation, the simulation setup reads $N=100$, $n=10$, $N_E=1000$, $N_i=10$, and $N_{ph}=10$. According to the analysis above, the numerical complexity of standard NEGF calculation in the typical effective mass situation is $O(10^{10})$ floating point operations, and the memory usage is $O(10^7)$ floating point

numbers; the numerical complexity of LRA approximated NEGF calculation is $O(10^7)$ floating point operations, and the memory usage is $O(10^5)$ floating point numbers. This observation demonstrates clearly that LRA method can reduce both numerical cost and memory usage significantly.

The comparison of the amount of floating point operations and memory usage between the exact and the approximated LRA approach illustrate that the LRA method offers approximated solutions of the NEGF equations much faster and with a much smaller memory load than the exact solutions. In fact, one can easily find NEGF equations of state-of-the-art devices that are only solvable when the LRA method is applied.

4.2.4. Benchmark and application

First the LRA method is benchmarked on a 50nm thick, homogeneous layer of GaAs with an effective mass of $m^*=0.067m_0$. The NEGF equations are discretized with a 0.5nm mesh spacing. The Fermi energies in both contacts are assumed to agree with the respective conduction band edge. The temperature is set to 300K. The conduction band in the device is set to be constant in the first and last 5nm of the device and to drop linearly by the amount of the applied bias voltage in the central 40nm of the device.

Fig. 4.1 shows the spatially resolved current density that results from an exact NEGF calculation as well as current densities of LRA calculations when the matrix rank is reduced to 20% and 10% of the original space. The exact calculation yields a spatially constant current in the device, since inelastic phonon scattering is included through a converged self-consistent Born approximation. At the device boundaries, the matrix elements of the exact contact self-energy are non-zero as is common in the NEGF method. This non-vanishing self-energy allows electrons to enter and leave the device, thus this contact self-energy violates current conservation at the device boundaries. In the LRA method, the contact self-energies are transformed into dense matrices. Their largest elements are still located close to the device boundaries, which causes the largest current fluctuations at the boundaries. The larger the matrix rank reduction is, the larger contact self-energy matrix elements within the device are. Consequently, the larger the rank reduction is, the higher the maximum amplitude of current density fluctuations within the

device is. The smaller the rank of the reduced real space is, the denser the contact self-energies are. This allows electrons to leave/enter the device at/to any device point in the reduced rank space. The non-constant current densities in Fig. 4.1 in the original real space indicate this kind of violation of particle conservation. Similar to the current fluctuations the density deviates from the exact solution stronger if the rank of the NEGF equations is reduced more. Fig. 4.2 shows the electron density in the homogenous layer of GaAs in equilibrium and at finite applied bias voltage. In both cases, the deviations are strongest close to the contacts.

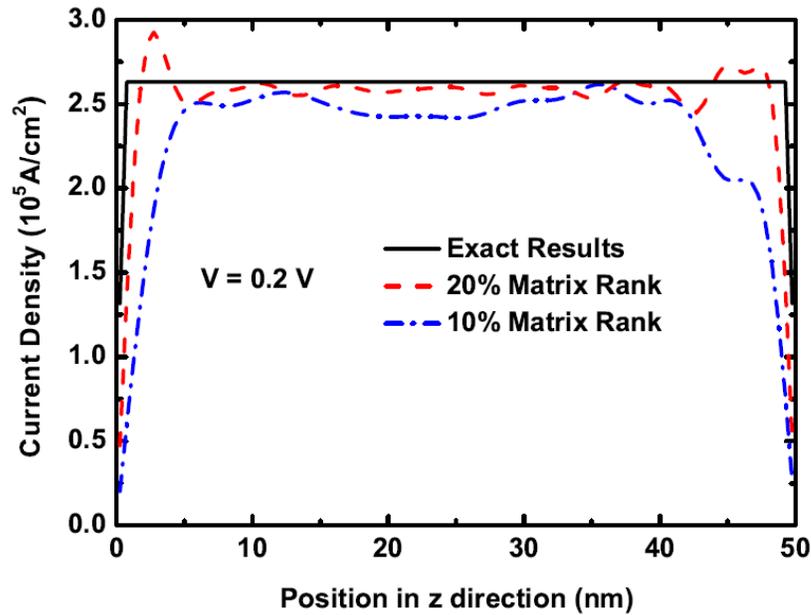


Fig. 4.1 Spatially resolved current density in the homogeneous structure with a linear potential drop of 0.2 eV. The lines show that the result calculated approximately with a reduction of the matrix rank down to 20% (dashed) agree well with the NEGF method solved exactly (solid), and the result with a reduction of the matrix rank down to 10% (dash-dotted) shows a larger deviation from the exact solution.

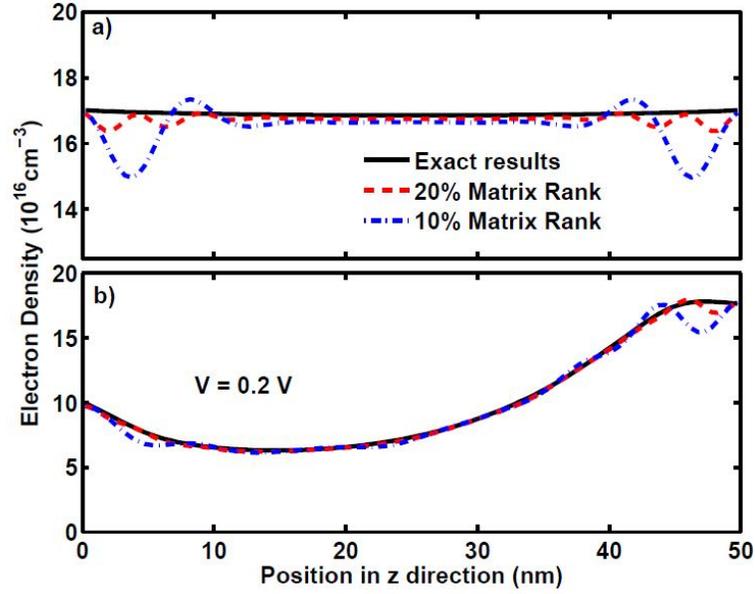


Fig. 4.2 Calculated electron density of the homogeneous device of Fig. 4.1 in (a) equilibrium and (b) when a linear potential drop of 0.2 eV is assumed. Results with approximated NEGF calculations where the matrix rank is reduced down to 20% (dashed) agree well with the exact NEGF calculations (solid). Results with matrix rank reduction down to 10% (dash-dotted) show larger deviations from the exact solution.

Both of the above figures indicate that the LRA method can reproduce exact NEGF results in the device center away from the contacts. Thus the current is averaged away from the contacts. All remaining current densities in this chapter are such device averaged results.

Fig. 4.3 shows I - V characteristics of the 50 nm thick homogeneous GaAs layer that have been calculated in the exact NEGF method, as well as in the approximated LRA method with various rank reduction levels. In addition, Fig. 4.3 also shows an exact NEGF calculation of the same device with a ten times coarser grid mesh. Similar to the previous figures, the larger deviation between the LRA results and the exact NEGF results is, the larger the rank reduction is. Nevertheless, a reduction of rank down to 10% is still able to well reproduce the I - V characteristics, since only a small fraction of electronic states in the low energy range contributes to the current density. In contrast, results of exact NEGF calculations with a ten times coarser grid deviate significantly from the full rank result. This is expected because a coarser real space yields a different numerical electron dispersion that deviates from the parabolic dispersion at the energies of interest.

It is worth to mention that the boundary conditions for the electronic wave functions of (4.5) are relevant for the efficiency of the LRA method. In agreement to similar findings of Mamaluy *et al.* [42][43], basis functions with Dirichlet boundary conditions turned out to be inferior to Neumann conditions. This boundary condition issue will be further discussed in Appendix A.

The LRA method and standard NEGF calculations were implemented and tested in MATLAB with 8 cores parallelization and are also available in NEMO5. For this homogeneous structure calculation, the measured computational time for matrix rank reductions down to 10% and 20% is reduced by factors of 35x and 87x respectively compared to the full solutions, as listed in Table 4.2. In the non-optimized LRA MATLAB implementation, most of the time is spent in transforming Green's functions between different basis representations. If further optimizations on matrix transformation are made, the factor of speed up can be even larger.

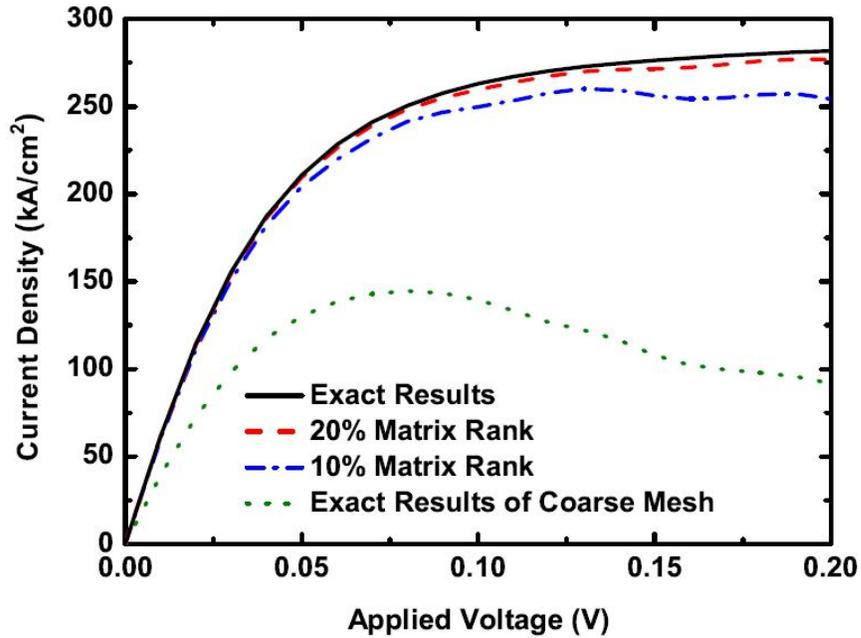


Fig. 4.3 Comparison of I - V characteristic of the structure of Fig. 4.1 shows that the approximated NEGF solutions where the NEGF matrix rank is reduced to 20% (dashed) agree well with the exact NEGF calculations (solid). The results with matrix reduction down to 10% (dash-dotted) show larger deviations. Results for the NEGF calculations with a 10 times coarser real space mesh (dotted) are shown to deviate significantly from the dense mesh case due to its different effective electron dispersion.

Table 4.2 The measured time consumptions for the three examples with matrix rank reductions down to 10% and 20% as well as the exact NEGF solutions are listed. The LRA method and the standard NEGF calculations were implemented in MATLAB with 8 cores parallelization. The results show significant speed ups of the LRA method.

	10% Matrix Rank	20% Matrix Rank	Exact Solution
50nm resistor	42s/87x	104s/35x	3658s/1x
RTD structure	287s/150x	1107s/39x	43058s/1x
1000nm resistor	Too aggressive	459hr	Memory exceeded

Next we explore the compatibility of the LRA method in quantum confined systems. The NEGF equations are solved in an 80nm GaAs/Al_{0.3}Ga_{0.7}As resonant tunneling diode (RTD) structure at 100K. The RTD consists of two 3nm wide Al_{0.3}Ga_{0.7}As barriers and a 5nm quantum well in the center. In addition, a 40nm flat band region is located at the emitter region. The effective mass for GaAs is $0.067m_0$ and $0.0919m_0$ for Al_{0.3}Ga_{0.7}As. The band offset between these two materials is 230meV. In the original real space representation, the device is discretized with a grid spacing of 0.5nm. The Fermi energies in the contacts are set to be 0.005meV beneath the respective conduction bandedges. The potential profile is assumed to be constant in the left most 40nm and to drop linearly in the remaining RTD region. This is illustrated by the solid line in Fig. 4.4 which shows the same conduction band profile of the RTD in an exact NEGF calculation and a 10% LRA approximated NEGF calculation. Fig. 4.4 also shows contour graphs of the energy and spatially resolved electron density of the RTD at vanishing in-plane momentum ($k_{||}=0$) when a voltage of 0.1375V is applied. Both results agree well: Fig. 4.4(b) deviates from (a) only at the energy of about 0.07 eV and positions $z \in (65,80)$. Even the confined state in the triangular quantum well locating at the left of the first RTD barrier (at energy of about 0.05eV and position 50nm) is well reproduced in the LRA calculation. This is remarkable, since electrons can enter this state effectively only via inelastic scattering. Therefore, inelastic scattering and tunneling are well reproduced with the LRA method. That can also be seen in Fig. 4.5, which shows the I - V characteristics of this RTD structure calculated in the exact NEGF method, as well as in the LRA method with 10% and 3.1% of the original matrix rank. Neither the current amplitude nor the resonance

value gets significantly altered when the NEGF equations are solved with only 10% of the original matrix rank. If the matrix rank is reduced too much, electronic states that are relevant for the transport are neglected. Consequently, the current density starts to deviate. This is illustrated in Fig. 4.5 with the I - V characteristic of a LRA approximated NEGF calculation of only 3.1% of the original matrix rank. The ratio of the matrix rank reduction can be estimated from the energy interval in which the states are occupied. For this RTD example, the measured computational time for matrix rank reductions down to 20% and 10% is reduced by factors of 39x and 150x respectively compared to the full solutions, as listed in Table 4.2.

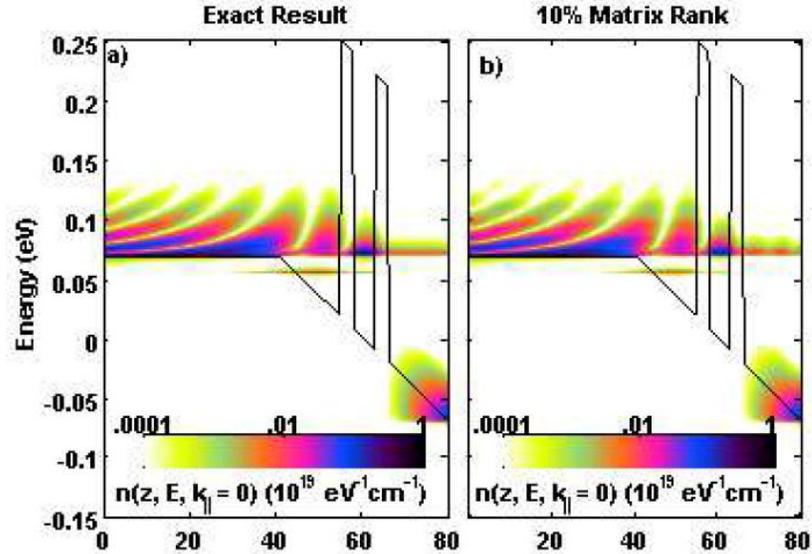


Fig. 4.4 Conduction band profile (solid line) and contour plot of the energy resolved electron density of the RTD structure described in the main text. The energy resolved density is calculated within the NEGF method (a) exactly and (b) approximately by a reduction of the NEGF equations' matrix rank to 10%. The results show that the filling of bound states in the triangular well by inelastic acoustic phonon scattering is well captured in the approximated method.

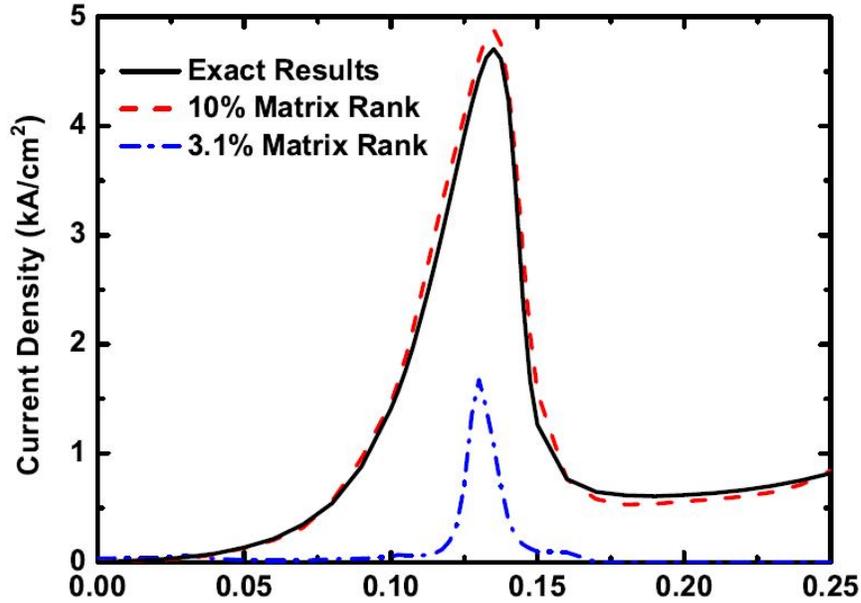


Fig. 4.5 I - V characteristic of the RTD structure of Fig. 4.4 calculated approximately with 10% (dashed) of the original matrix rank agrees well with the exact solution (solid). However, too aggressive matrix rank reduction of 3.1% (dash-dotted) fails to reproduce the results due to the neglecting of relevant electronic states.

The new LRA method is not only more efficient than the standard NEGF approach; it also opens up a space of device configurations that previously could not be tackled. The following considers electronic transport in the presence of inelastic phonon scattering in a 1000nm long homogeneous GaAs layer. In the range of 100nm within the source and the drain contact/device interface, the conduction band is assumed to be constant. In the remaining device, the conduction band drops linearly according to the applied bias voltage. The temperature of the phonon bath and the electrons in the contacts is 300K and the Fermi levels of the contacts are set to be 0.1eV above the respective conduction band edge. The system is originally discretized with a mesh spacing of 1nm. The resulting NEGF equations are approximated with a 20% matrix rank. This reduces the numerical complexity of the NEGF equations such that they have been solved on a single CPU without recursive approaches. The nature of the transport is tuned from purely ballistic to almost drift-diffusion like by increasing the deformation potential D_{ac} of the phonon scattering self-energy of (4.3). Hereby, three different scattering potentials have been

considered: 27eV, 60eV and 135eV, which correspond to scattering rates of 10^{12}s^{-1} , $5\cdot 10^{12}\text{s}^{-1}$ and $2.5\cdot 10^{13}\text{s}^{-1}$ for electrons with kinetic energy of 0.3eV. The impact of the scattering is illustrated in Fig. 4.6 as it shows the calculated I - V characteristics of the device with various scattering strengths. The I - V characteristics are almost ohmic in the case of a deformation potential of 135eV. The nature of transport at this large deformation potential can be understood from Fig. 4.7. It shows the energy and spatially resolved electron density for the 1000nm long resistor in the case of 0.1V applied bias voltage. Electrons that originate from the source contact propagate about 400nm in the device before they start to significantly dissipate energy. Then, however, these electrons follow the potential drop of the device and thereby start to maintain a local equilibrium distribution. In this way, the electrons experience a transition from effectively ballistic transport into the drift-diffusion of the rightmost 500nm of the device.

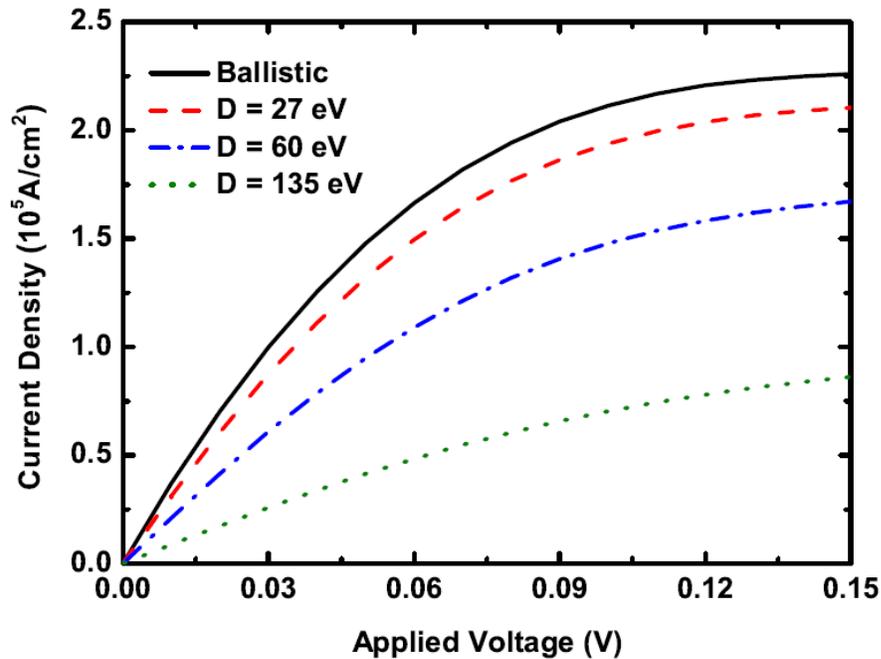


Fig. 4.6 The I - V characteristic of the 1000nm homogeneous structure described in the main text when different values for the deformation potential are used: 27eV (dashed), 60eV (dash-dotted), and 135eV (dotted), and ballistic results are shown as a solid curve.

All results are determined from approximated NEGF equations with a matrix rank reduction down to 20% of the original rank. With the deformation potential of 135eV the I - V characteristic is almost ohmic.

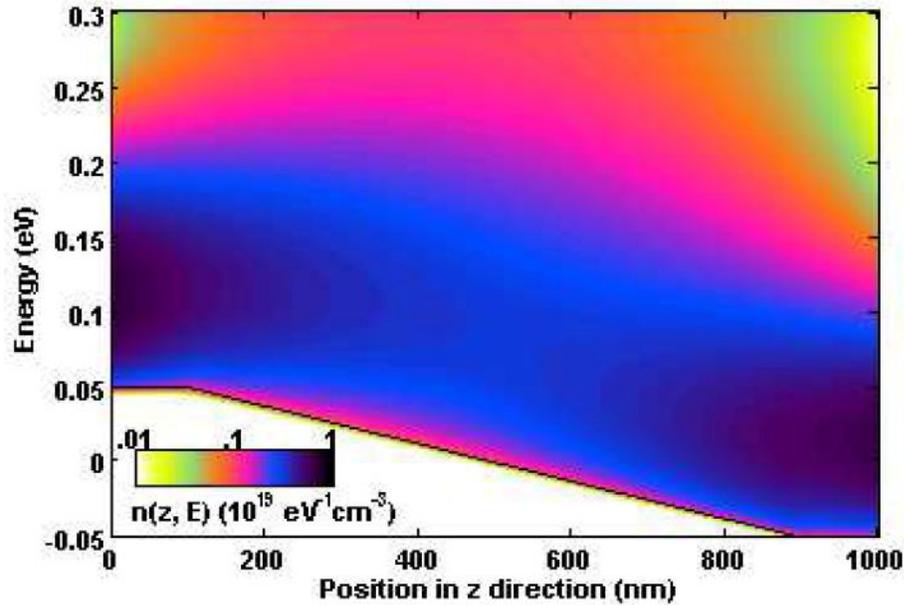


Fig. 4.7 Energy resolved electron density of the 1000nm long structure of Fig. 4.6 with a deformation potential of 135eV and a potential drop of 0.1V. Energy dissipation happens around the center of the device suggesting a transition from effectively ballistic transport into the drift-diffusion of the rightmost 500nm of the device.

4.2.5. Summary

The LRA method is applied to efficiently and accurately solve the approximated NEGF equations in the effective mass approximation. It is shown that the method reliably solves the electronic transport in the ballistic and incoherently scattered transport regime. Quantum effects that are natively included in the NEGF equations are accurately reproduced by the LRA method, but with a smaller numerical load. The method is illustrated in homogeneous resistors and RTDs and shows very good agreement with the exact NEGF calculations. The method is applied to a 1000nm long GaAs resistor with incoherent scattering to show the efficiency and power of the method.

4.3. LRA in Tight Binding

4.3.1. Basis function

In the above section the objective and basic concept of the LRA method were discussed. In the TB case, given the extra orbital degrees of freedom to capture the many band effects, the numerical load is much higher than the effective mass case. As a result, the LRA method is very promising in the NEGF calculations with TB representations. The same idea holds in the TB case: only a few states around the chemical potential are occupied and relevant to transport problems, thus one can find a basis function that can represent those relevant states, and transform the original equations into the reduced space. The key step of this method is to find the proper basis function to transform the original NEGF equations into the reduced space. In the effective mass case, it is already shown that the eigenstates solved from the free particle Hamiltonian with the Neumann boundary condition in (4.5) can serve as good basis functions. Further efforts can be made to generate a reduced real space if inelastic scattering exists. In the TB case, it is not that trivial, since the Hamiltonian with Neumann boundary condition cannot perfectly represent the quasi-particle with different injected energies. Instead, an open boundary Hamiltonian is used to generate the basis functions, which is written as

$$[H_0(k_{\parallel}) + \Sigma^R(k_{\parallel}, \varepsilon)]\phi_i = E_i\phi_i, \quad i = 1, 2, \dots, n \quad (4.9)$$

where H_0 is the device Hamiltonian, Σ^R are self-energies from all contacts, and k_{\parallel} denote the in-plane momentums. Only the orthonormal eigenstates $\{\phi_i\}$ with E_i closest to the considered particle energy ε are considered. Notice that the self-energy is energy ε and momentum k_{\parallel} dependent, so are the eigenstates $\{\phi_i\}$. The LRA method can be applied in the self-energy calculations also, which will be discussed in chapter 4.4. Here we talk about how to solve for the Green's functions in the device. Once the eigenstates are obtained, the next step is to construct the $N \times n$ transformation matrix P with these eigenstates, at each energy ε and momentum k_{\parallel} . After the transformation matrix is obtained, the NEGF equations are transformed into the reduced space

$$\begin{aligned}
G_P^R &= \{P^\dagger[\varepsilon - H_0(k_{||}) - \Sigma^R(k_{||}, \varepsilon)]^{-1}P\}^{-1} \\
\Sigma_P^R &= P^\dagger \Sigma^R(k_{||}, \varepsilon)P \\
G_P^< &= G_P^R \Sigma_P^< G_P^{R\dagger}
\end{aligned} \tag{4.10}$$

Transmission can be solved in the reduced basis P following the standard transmission equation [1]. To solve charge density is a little tricky. After $G_P^<$ at each energy ε and momentum $k_{||}$ is obtain in the reduced basis, an integration over all relevant ε and $k_{||}$ is needed to calculate the charge density. Since the basis functions are ε and $k_{||}$ dependent, this integration cannot be done in basis P . There are two ways to solve for the charge density. The first way is to transform $G_P^<$ of each ε and $k_{||}$ back into real space.

$$G^< = P G_P^< P^\dagger \tag{4.11}$$

Since only the diagonal of $G^<$ is needed for charge density, this transformation (which is matrix-matrix products) can be done column-wise and only solve for the diagonal elements of $G^<$. Therefore, the memory and time consumption is negligible. After the transformation of (4.11) is done for each ε and $k_{||}$, an integration over $G^<$ is performed to obtain the density. The second way to solve for the density is to find a common basis in the reduced space P . Once a common basis is found, one can perform the integration in the reduced space and transform the final results back into the original space to obtain the density. The methods to find a common basis will be discussed in chapter 4.3.3. Here, all the density results are calculated with the first method.

4.3.2. Benchmark and application

The LRA method is demonstrated in a squared Si nanowire with 5nm diameter in $sp^3d^5s^*$ TB basis. A short device with only one unit cell along the transport direction is used here such that the exact NEGF solution can be easily computed. A homogeneous electrostatic potential in the whole device is applied. Fig. 4.8 shows the relative deviation of the electron density in the cross-section of the Si nanowire. The relative deviation defines as:

$$dev = 100\% \times |exact - LRA| / |exact| \tag{4.12}$$

The LRA is done using (4.10) and (4.11), with approximated matrix rank of 10% of the original rank. The maximum deviation is around 1.5%. This suggests that with NEGF equations solved in only 10% of the original matrix rank, the density is well reproduced. The density along the center of the nanowire cross-section is shown in Fig. 4.9. The inset of Fig. 4.9 shows that the LRA result deviates from the exact solution only slightly. It is interesting to notice that the electron density does not show an oscillation close to the boundary as seen in Fig. 4.2. This indicates that the basis functions generated with (4.9) represent the quasi-particles of the system perfectly.

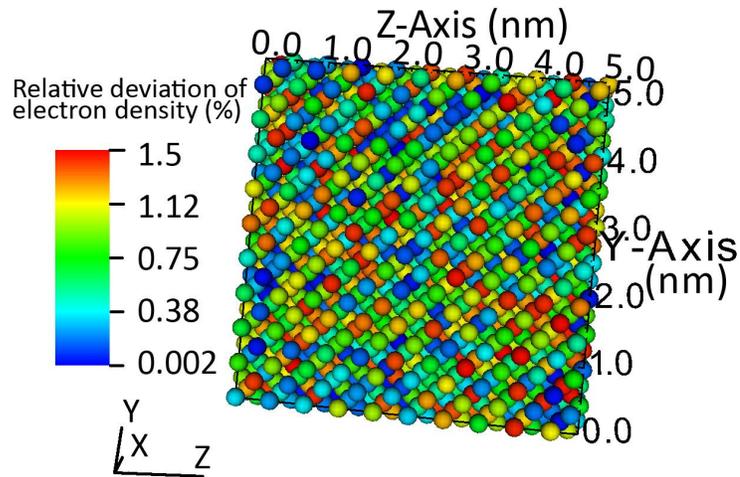


Fig. 4.8 Relative deviation of electron density in a squared, 5nm diameter Si nanowire of exact NEGF calculation and of NEGF calculations approximated with 10% of the original matrix rank. Transport is along the x-axis. The small relative deviations suggest that with 10% of the original matrix rank the electron density is nicely reproduced.

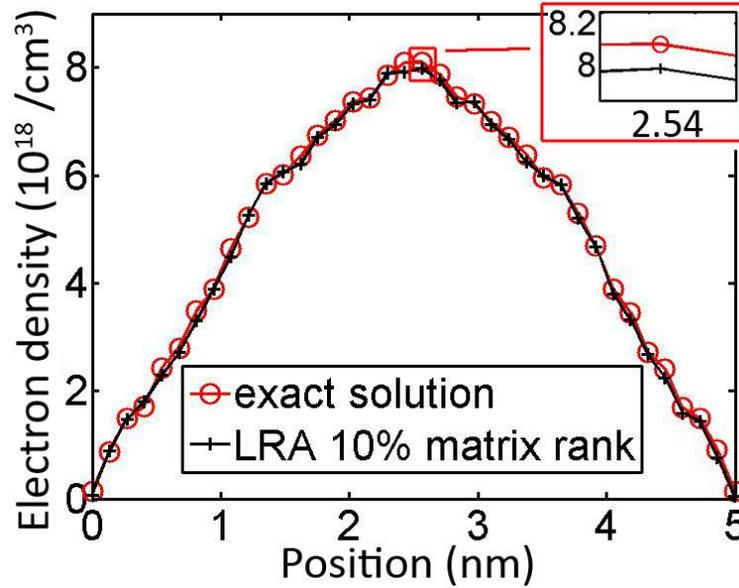


Fig. 4.9 Electron density along the center of the nanowire shown in Fig. 4.8, of the exact NEGF calculation (circle) and of NEGF calculations with 10% of the original matrix rank. The small deviations suggest that with 10% of the original matrix rank the electron density is well reproduced.

The transmissions for the same structure are shown in Fig. 4.10 (valence band) and Fig. 4.11 (conduction band). It is shown that with only 10% of the original matrix rank, the LRA method perfectly reproduces the transmissions up to $\sim 0.7\text{eV}$ above the conduction band edges and below the valence band edges. The LRA results of 2% of the original matrix rank are also shown. Similar to the effective mass case, with aggressive reductions the deviations between the LRA results and the exact solutions become larger. However, the transmissions can still be reproduced up to $\sim 0.25\text{eV}$ ($10k_B T$) above the conduction band edges and below the valence band edges with only 2% of the original matrix rank.

All of the calculations are done in NEMO5. With LRA calculations of 10% of the original matrix rank, an overall speed up of 8x is observed in these calculations. This is effectively limited by the solution of eigenvalue problem in (4.9). The selection of eigenvalue algorithms and solvers will be briefly discussed in chapter 4.5.

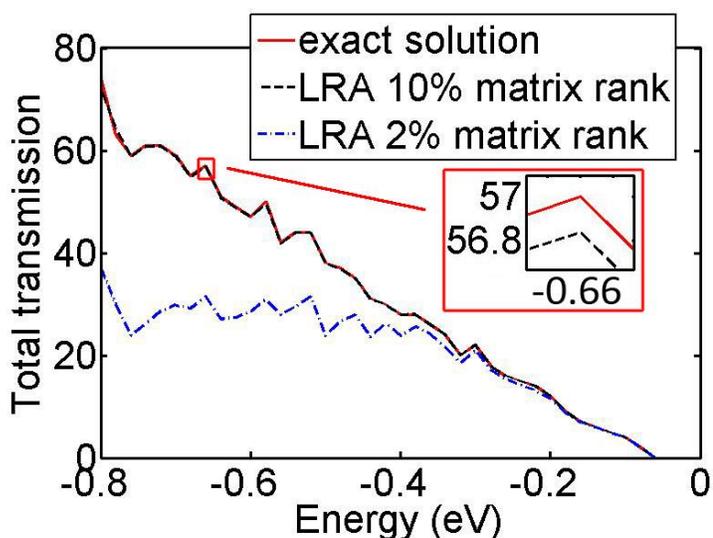


Fig. 4.10 Total transmissions of the exact NEGF calculation (solid) and of NEGF calculations with reduced matrix rank for the valence band of the Si nanowire shown in Fig. 4.8. With 10% of the original matrix rank the transmission is well reproduced, while too aggressive reduction of 2% of the original rank leads to larger deviations.

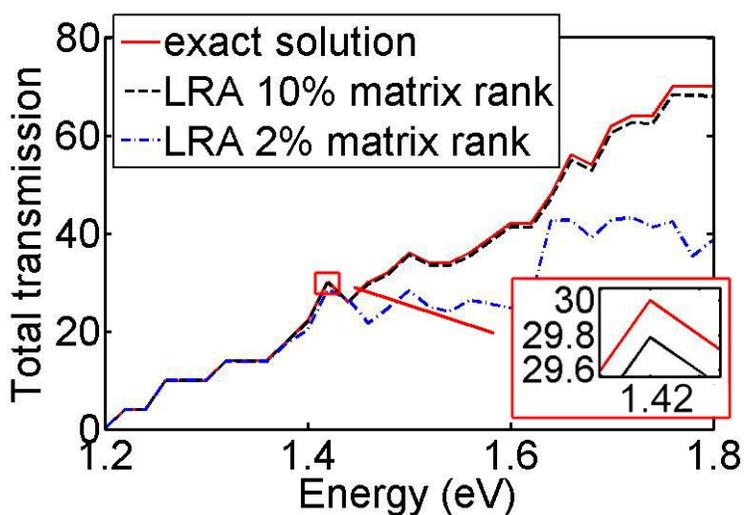


Fig. 4.11 Total transmissions of the exact NEGF calculation (solid) and of NEGF calculations with reduced matrix rank for the conduction band of the Si nanowire shown in Fig. 4.8. With 10% of the original matrix rank the transmission is well reproduced, while too aggressive reduction of 2% of the original rank leads to larger deviations.

The second benchmark case is performed on a 3nm diameter, 8nm long Si nanowire in $sp^3d^5s^*$ TB model, with 0.1V source-drain bias and 0.1eV electrostatic potential drop

from source to drain. An exact solution is not feasible in direct NEGF calculation so that RGF is used in the exact solution. Fig. 4.12 shows the transmission calculated with RGF and with NEGF calculations of 10% of the original matrix rank. The results show that the basis of (4.9) works in non-flat potential case, where the modes of different energy points couple to each others.

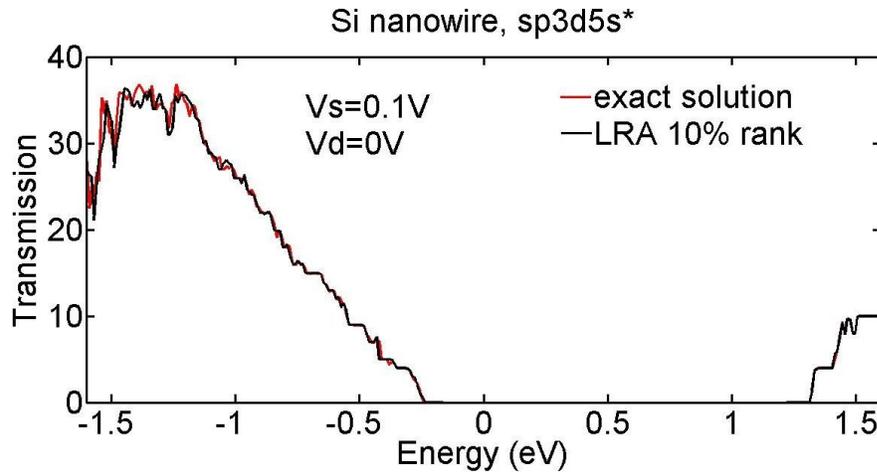


Fig. 4.12 Total transmission of a 3nm diameter, 8nm long Si nanowire with 0.1eV linear potential drop is solved. It shows that NEGF calculations with 10% of the original matrix rank (black) nicely reproduce the transmission of the exact NEGF calculation (red).

The third case is for a tunneling field effect transistor (TFET) with a non-trivial transport path. The TFET is a broken gap structure, whose transport is based on mixture of electrons and holes. The structure of the TFET is shown in Fig. 4.13. The source region of the TFET is constructed with 10nm×10nm GaSb, and the drain region is constructed with 10nm×20nm InAs. Periodic boundary conditions are applied in the transverse direction. The transport path is from GaSb to InAs, which is a non-linear 2D path. Fig. 4.14 shows the transmission calculated with exact NEGF and with LRA of 5% of the original matrix rank. It is shown that the results are well reproduced with LRA in both valence band, conduction band, and the broken gap range where tunneling happens. This suggests that the LRA method can work perfectly for non-trivial physical situations with non-trivial transistor geometries.

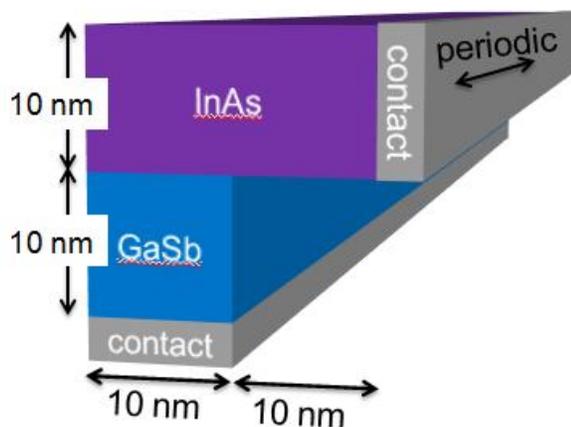


Fig. 4.13 L-shaped GaSb-InAs TFET transistor.

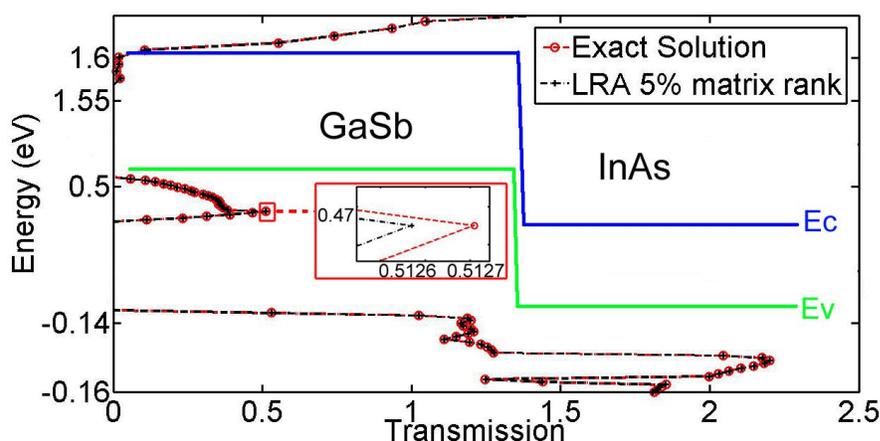


Fig. 4.14 The total transmission for the GaSb-InAs TFET shown in Fig. 4.13, with exact NEGF calculation (red) and approximated NEGF calculation of 5% of the original matrix rank (black). The band edges of the GaSb and InAs are shown to indicate the energy range for tunneling. The results are well reproduced with LRA in both valence band, conduction band, and the broken gap range where tunneling happens.

The last example is a $12\text{nm} \times 12\text{nm}$ squared Si nanowire represented in 10 band TB model. The exact solution of this structure is not feasible without a supercomputer since the memory consumption is huge. The LRA method with 10% of the original matrix rank is applied in this structure and the transmission is solved with ~ 300 processing cores in ~ 35 hrs. The transmission results are shown in Fig. 4.15 (valence band) and Fig. 4.16 (conduction band). This calculation demonstrates the power of the LRA method.

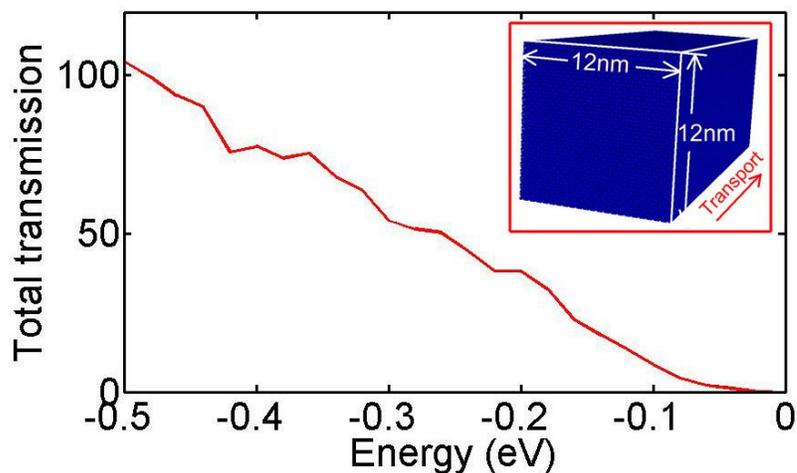


Fig. 4.15 Total transmission calculated with 10% of the original matrix rank for the valence band of a 12nm diameter Si nanowire without a supercomputer. The inset shows the schematic of the nanowire. This calculation shows the power of the LRA method.

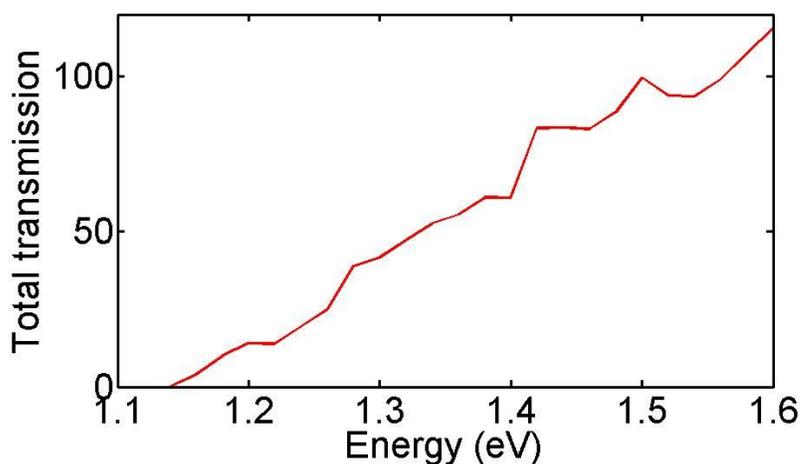


Fig. 4.16 Total transmission calculated with 10% of the original matrix rank for the conduction band of a 12nm diameter Si nanowire without a supercomputer. This calculation shows the power of the LRA method.

4.3.3. Common basis in LRA-TB

It is more convenient to have a common basis, rather than an ε and k_{\parallel} dependent one, since an integration over different energy and momentum is usually needed, such like the calculation of charge density, or the case with inelastic scattering. In this chapter, several methods to find a common basis is introduced, and their advantages and disadvantages are discussed.

The simplest and most straightforward method is to use the basis functions obtained from (4.9), and apply them in a small ε and k_{\parallel} interval. As an example, a 12nm long GaAs quasi-1D structure in $sp^3d^5s^*$ TB model is simulated. Here, for simplicity the calculation is done for $k_{\parallel}=0$ only. Fig. 4.17 shows the transmissions for the valence band and conduction band, with exact NEGF calculations and with LRA. The basis functions are solved using (4.9), but only in several specific energy points. For the valence band, the basis functions are solved in -1.9eV and -0.1eV, and applied in the energy range below and above 1eV respectively. For the conduction band, the basis functions are solved in 1.45eV and 2.9eV, and applied in the energy range below and above 2.25eV respectively. Fig. 4.17 shows that the results are reproduced only very close to the energy points where the basis functions are generated. For the energies far away, the transmissions deviate significantly from the exact solution. This suggests that the straightforward method can work only in a very tiny energy interval but it is easy to apply and does not require much extra effort.

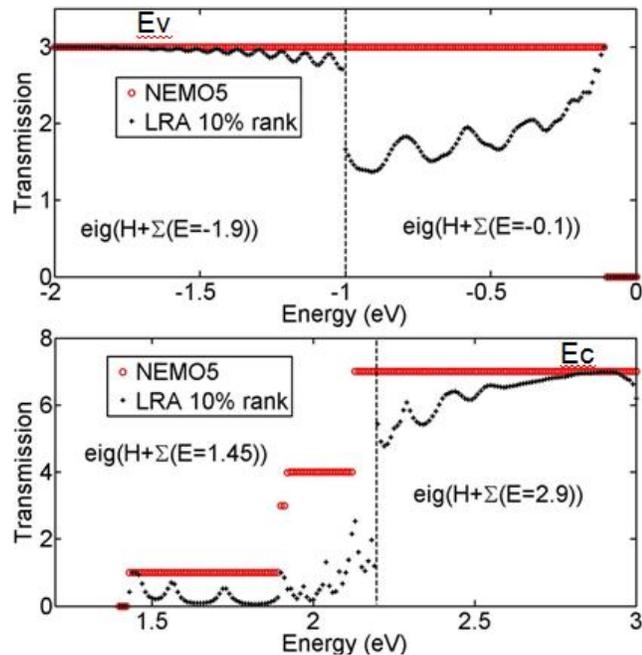


Fig. 4.17 Transmissions of the valence band and conduction band for a 12nm long quasi-1D GaAs structure. Calculations are done exactly and with LRA of 10% of the original rank. The basis functions are generated in a few specific energy points only. In the LRA case, the transmission is reproduced only in a tiny interval of the energy points where the basis functions are generated.

An improved method can be found by looking into the correlations of the basis functions. The basis functions from two different energy points are not totally independent, since their corresponding eigenstates are connected within the complex band structures of the system. Therefore, the idea to improve the common basis obtained above is to find out the correlations of the two basis sets, and generate a new basis set with this correlation information. First step is to generate a new matrix with the basis sets from two different energy points. Assuming the transformation matrices from the two basis sets are P_1 and P_2 , a new matrix $P=[P_1,P_2]$ combining all eigenvectors in P_1 and P_2 is formed, and the eigenvalue problem is performed as below

$$P^\dagger P \phi_i = E_i \phi_i, \quad i = 1, 2, \dots, n \quad (4.13)$$

The eigenvalues E_i contains the correlations of all the vectors in the P matrix, or effectively the correlations of all the eigenvectors in P_1 and P_2 . The n largest values of E_i which represent the most related portions of P_1 and P_2 as well as the corresponding eigenfunctions $\{\phi_i\}$ are used to generate the new transformation matrix T . The NEGF equations are solved in the same way as in (4.10), but with the transformation matrix T instead.

Fig. 4.18 shows the transmissions for the same structure as in Fig. 4.17, but with the new method described above. The same four energy points are used to generate P_1 and P_2 matrices. The transmission results are perfectly reproduced for both the valence band and conduction band with the method described in (4.13). This physically makes sense since the most relevant information of the valence band and conduction band are given by (4.13) in the transformation matrix. The numerical effort is the additional eigenvalue problem in (4.13). This is usually not relevant since n is much smaller than the original matrix rank N .

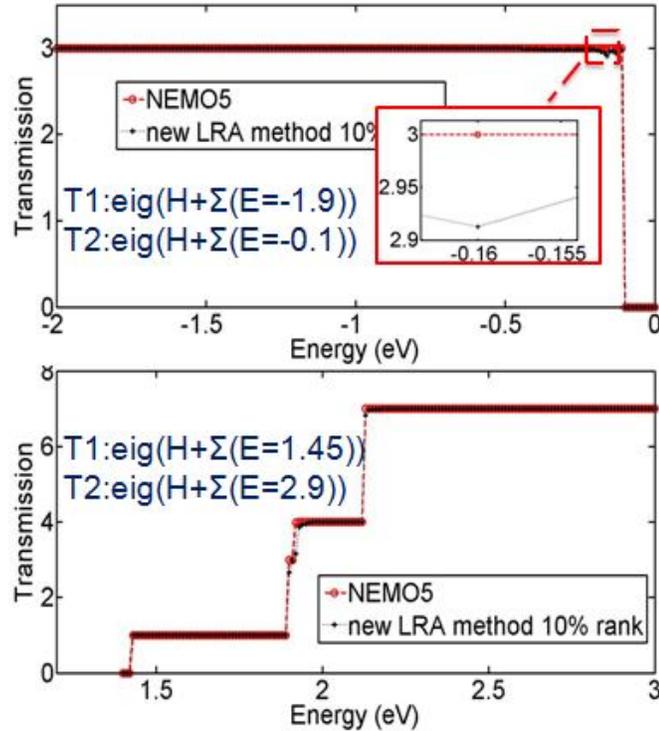


Fig. 4.18 Transmissions of the valence band and conduction band for a 12nm long quasi-1D GaAs structure. Calculations with LRA of 10% of the original matrix rank agree well with the exact calculations. The basis sets are generated following (4.13). The results suggest that the new method of (4.13) is a potential solution in finding a common basis.

4.3.4. Summary

In summary, the LRA method is successfully applied in NEGF within the TB representation. The method is demonstrated on homogeneous structures with and without an electrostatic potential involved. It is also shown to be applicable to complicated physical situations like TFETs with non-linear transport geometries. The transmission for a 12nm diameter nanowire in 10 band TB model is shown to be solvable without a supercomputer using LRA. Some improved methods in finding the basis functions in TB case are also discussed.

4.4. LRA in Contact Self-energy

To solve for the contact self-energy in empirical TB representation is expensive, especially for nanowires with large diameters. An efficient solution of the contact self-energy is therefore required. The most efficient contact self-energy method so far is the

modified version of the transfer matrix method [6]. In this method, a normal eigenvalue problem is solved and the contact modes are obtained. Since in realistic devices usually only the propagating modes are relevant, the concept of LRA can be applied here to reduce the computational effort. As shown in chapter 3.4, the heavy numerical loads of the transfer matrix method are the eigenvalue problem and a couple of matrix-vector, matrix-matrix products. The idea of LRA is to solve for the propagating modes only (or include some slowly decaying modes also) to reduce the effort in solving the eigenvalues and the matrix products.

A detailed analysis of the eigenvalue problem in (2.18) shows that the eigenvalues have some very special spectrum. Assuming the eigenvalue of (2.18) is λ , one can derive

$$\lambda = \frac{1}{e^{-ik\Delta} - 1} \Rightarrow e^{-ik\Delta} = 1 + \frac{1}{\lambda} \quad (4.14)$$

Note that the important modes for transport problem are the propagating ones, whose wave vectors k are real numbers. Therefore, it can be shown that

$$\begin{aligned} 1 &= |e^{-ik\Delta}| = \left| 1 + \frac{1}{\lambda} \right| = \left| \frac{\lambda + 1}{\lambda} \right| \\ &\xrightarrow{\lambda = a + bi} \\ 1 &= \left| \frac{(a+1) + bi}{a + bi} \right| = \frac{\sqrt{(a+1)^2 + b^2}}{\sqrt{a^2 + b^2}} \\ &\rightarrow a = -0.5 \end{aligned} \quad (4.15)$$

So the propagating modes are corresponding to the eigenvalues with real part of -0.5eV, and the decaying modes are decaying faster away from -0.5eV as shown in Fig. 4.19. Ideally only the propagating modes are relevant for transport, thus in the LRA method only the spectrum close to the real part of -0.5eV are solved in (2.18).

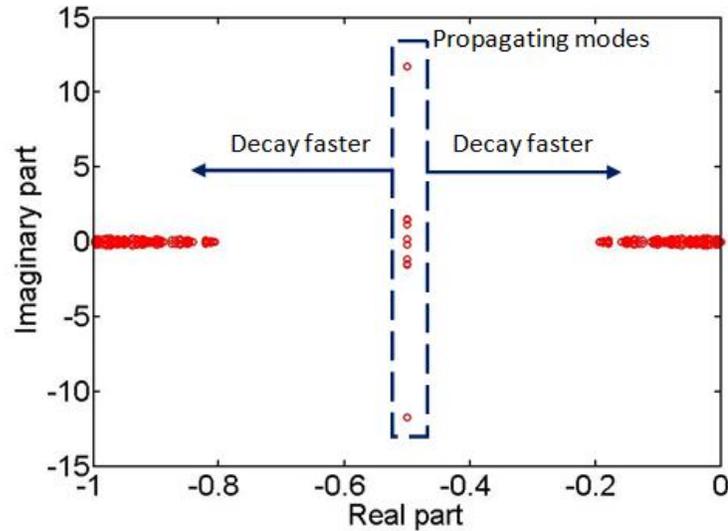


Fig. 4.19 Typical eigenvalue spectrum of (2.18). The propagating modes are corresponding to the real part of -0.5eV , and the decaying modes are decaying faster away from that.

Since the number of propagating modes is usually very small compared to the whole eigenvalue spectrum, the LRA method can significantly reduce the computational efforts in solving (2.18). Furthermore, all the following multiplications in (2.19) and (2.20) are speed up since only the eigenvectors of the propagating modes are relevant. An example of 3nm diameter, 20nm long Si nanowire in 10 band TB model is used to verify this method. The transmission and density with exact solution and with LRA of 10% of the original problem rank (including the propagating modes and some slowly decaying modes) are shown in Fig. 4.20 and Fig. 4.21. It is shown that with 10% LRA the transmission and density match well with the exact solution. A speed up of 3x in the contact self-energy calculation, and 30% in the total QTBM solution is achieved.

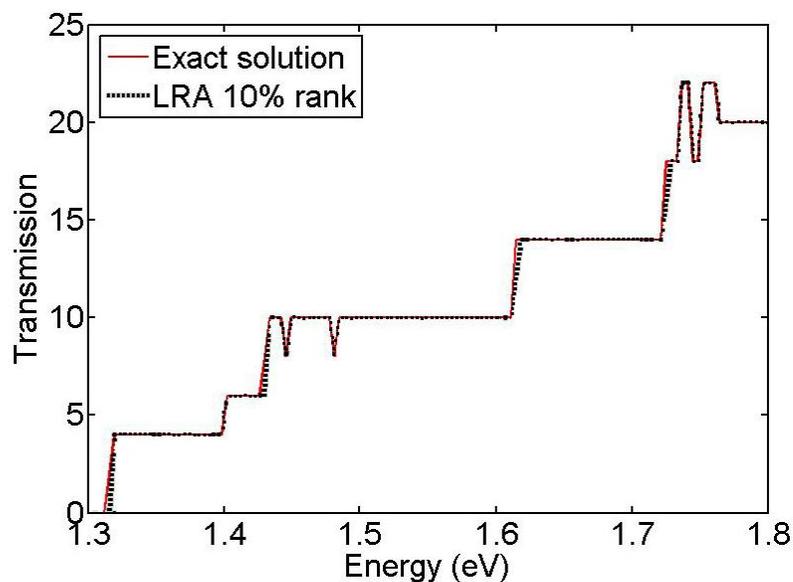


Fig. 4.20 Transmission solved with LRA of 10% of the original rank matches well with the exact solution.

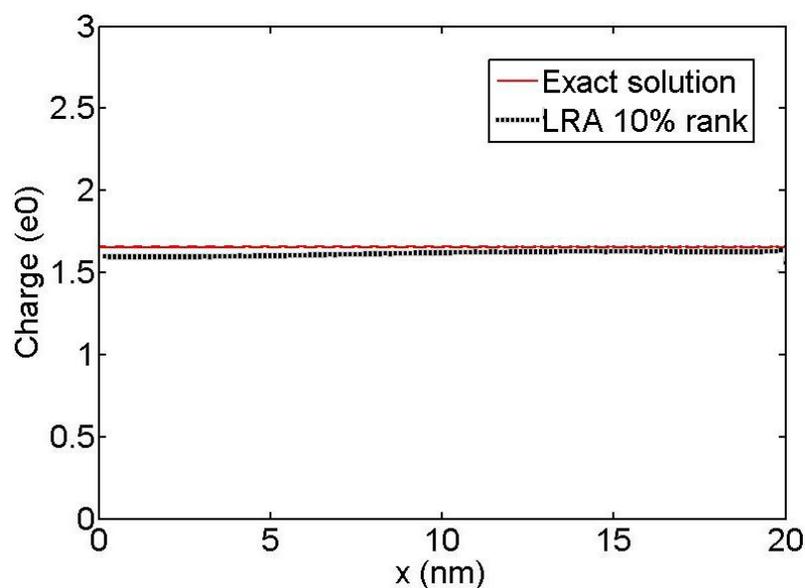


Fig. 4.21 Density solved with LRA of 10% of the original rank matches well with the exact solution.

It is worth to mention that this LRA method can work well for realistic devices with length $>10\text{nm}$. For tiny and short system ($<10\text{nm}$), the decaying modes become important since they can penetrate from the source contact to the drain contact before they vanish.

In this case, one will have to increase the LRA ratio to include more decaying modes in the eigenvalue problem in (2.18). Another problem is when there are electrostatic potentials coming into play, the modes will mix in the device channel so that one will have to increase the LRA ratio also.

4.5. The Eigenvalue Solvers

Since the efficiency of the LRA method depends strongly on the eigenvalue problem, this section will briefly discuss the choices of several eigenvalue solvers.

If all the eigenvalues are required, as in the normal solution of (2.18), and the LRA process like (4.7) and (4.13), one should use Lapack solvers (dgeev/zgeev) [39] since those are the most efficient ones. In NEMO5, the Lapack solver is used by default.

If only a portion (e.g. 10%) of the eigenvalues is solved, the Krylovschur solver [48] is a good choice. In the LRA method of the effective mass case (4.5), the TB case (4.9), and the self-energy case, the Krylovschur solver within PETSc/SLEPc library is used. There are several spectrum options in this solver, such as the largest/smallest magnitude, largest/smallest real part, largest/smallest imaginary part, or real part around some specific values, etc. It is recommended that for eigenvalue problems where smaller than 30% of all the eigenvalues are calculated, one can use Krylovschur solver [49]. Otherwise, one should use Lapack, since to solve over 30% of all the eigenvalues Krylovschur will be slower than Lapack.

There are other eigenvalue solvers, such as the FEAST solver [50], and the CISS solver [51], that are developed for solving eigenvalues within a complicated spectrum. These solvers can be applied to the LRA method in contact self-energy for the spectrum in (4.14), or even to the original generalized eigenvalue problem as shown in [6]. The feature of solving complicated spectrum for normal and generalized eigenvalue problem within these solvers can open up further explorations of the LRA method in the contact self-energy calculations.

4.6. Summary

In summary, a low rank approximation method is developed for NEGF calculation in both effective mass approximation and empirical TB representation, and the contact self-

energy calculation. Devices from simple homogeneous structures to RTDs and non-linear shape TFETs are studied using this new method. With only a few percent of the original matrix rank, the method can solve the transmission, current density and charge density for these devices accurately compared to the exact calculations, but with a significant speed up of several to a hundred times.

5. GENERAL LEAD ALGORITHM FOR CONTACT SELF-ENERGY

5.1. Introduction

As state-of-the-art semiconductor devices become smaller, the contacts gain relevance for the overall device performance. Details of the contact bandstructure and geometry influence the device transport characteristics, especially for low dimensional devices. Consequently, this issue motivates an increasing number of experimental and theoretical studies [52]-[54].

The NEGF method [1] is the standard approach to model open boundary devices in the nanometer length scale, where coherent quantum effects as well as incoherent scattering are present. Open boundary conditions of the device are taken into account by contact self-energies, which represent the charge injection and extraction effects of the contacts. The self-energies play a special role: they allow the assumption of specific charge distributions and density of states of the contact transport carriers. Previous algorithms that enable the injection of carriers from complex bandstructures rely on ideal contacts [55][56]. The distinction of contacts and central device allows limiting the explicit NEGF calculation to the central device area. This significantly reduces the numerical load [1]. There are well-known algorithms to solve self-energies for ideal contacts, such as the Sancho-Rubio method [10] and the transfer matrix method [6]. In the Sancho-Rubio method, the surface Green's function is iteratively solved and translated into a contact self-energy. In the transfer matrix method a generalized eigenvalue problem is solved to obtain the contact modes. These modes are then translated into the contact self-energy. However, both methods are based on the assumption of semi-infinite, periodic contacts. Realistic contacts are not ideal and often have non-linear geometries [54] or imperfections like disordered alloys or defects. Those effects cannot be captured by the two methods. Previous work introduced a lead algorithm based on the complex absorbing potential (CAP) [24]. So far, this method has been applied to ideal contacts [24]-[27]. In this chapter, a general lead method based on

the extension of the CAP method is developed to solve any sort of contact, be it regular (i.e. periodic as covered by previous algorithms) or irregular (i.e. including disorder effects or complicated geometries). After the algorithm is described, it is applied within the NEGF method on regular contacts to compare with the established transfer matrix method. Then the algorithm is applied to trumpet shaped contacts to prove its applicability to non-periodic contacts. Finally, results of a $\text{Si}_{0.5}\text{Ge}_{0.5}$ random alloy device show how important non-ideal contacts are for quantitative device performance predictions.

5.2. Algorithm Details

5.2.1. The contact structure

Unlike the Sancho-Rubio method or transfer matrix method, where only one unit cell is used in the contact calculation, the general lead algorithm considers explicitly the actual contact structure in the calculation. Realistic contacts are typically with the dimension of $>100\text{nm}$, making it numerically too expensive to solve for the surface Green's function. In practice, only a small portion of the actual contact structure closest to the device/contact interface is taken into the numerical solution. Such a contact portion is partitioned into minimal slabs, and then an iterative solution similar to the RGF [20] approach is performed in these slabs until it reaches the convergence of the surface Green's function. In this way, the length of the numerical contact portion considered in the general lead routine is determined by the convergence of the surface Green's function.

5.2.2. The iterative solution

Usually only the last few (~ 10) nanometers of the concrete contact structure that is closest to the device is considered in the calculation. This contact is partitioned into slabs and the surface Green's functions g_n for each slab are solved recursively following the directive iterative self-energy method of (2.15) until g_n is converged. To ensure the numerical efficiency, the contacts need to be partitioned in the way that the slabs are kept in a minimum size.

Equation (2.15) is a purely ballistic calculation, thus it is very sensitive to the initial guess. A bad initial guess will propagate towards the device interface ballistically and make it difficult to converge. One way to improve the convergence is to impose an artificial dephasing to the contact Hamiltonian to blur the initial guess. This dephasing has to change smoothly towards the device and vanish at the device/contact interface in order to make sure that the surface Green's function is ballistic and the DOS matches the device. This dephasing serves as a CAP as described in [24].

The dephasing function can be defined in different shapes. In fact dephasing functions with polynomial shape, logarithmic shape and exponential shape are tested and all show better convergence than pure ballistic solution as defined in equation (2.15). The exponential shape dephasing is adopted in this work since it shows the best convergence and its parameters are easier to tune. (2.15) is rewritten as

$$\begin{aligned} g_n &= (E + i\eta_n - H_{n,n} - \Sigma_n)^{-1} \\ \eta_n &= \eta_0 e^{-\lambda(n-1)} \\ \Sigma_n &= H_{n,n-1} g_{n-1} H_{n-1,n} \end{aligned} \quad (5.1)$$

Here, $H_{n,n}$ is the Hamiltonian of the slab n , and Σ_n is the corresponding contact self-energy. The first slab's contact self-energy vanishes ($\Sigma_0=0$). An imaginary damping potential η_n with exponential shape is chosen as the CAP. It decays with increasing n , i.e. it is the smaller the closer the respective contact slab is to the device. The initial damping potential η_0 is comparable with the energy E . The decay parameter λ is chosen such that η_n is negligible for large n . The purpose of η_n is to improve the convergence behavior, i.e. to minimize the contact length for a converged contact self-energy Σ_n . Depending on the actual contact structure and geometry, the optimal η_0 and λ can vary. Nevertheless, we have found that $\eta_0=-0.5\text{eV}$ and $\lambda=0.1$ provide good convergence for all considered device structures in this work. Fig. 5.1 shows the program flow of equation (5.1). It is important to notice that only the surface Green's function at the last slab g_N is needed for self-energy, such that only the last matrix block is stored. Actually, g_n only depends on Σ_{n-1} , so that one can release the memory for the $n-1$ 'th block once the n 'th block is solved. As a result, the peak memory of this algorithm only depends on the storage of a dense matrix

of one slab size. This memory slim feature makes the algorithm very appealing for realistic device simulation.

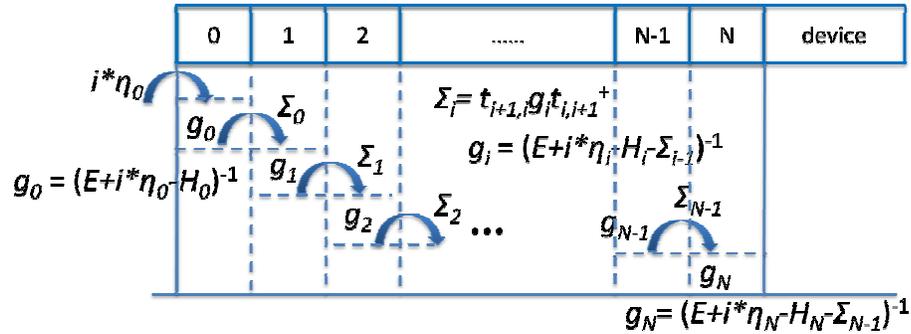


Fig. 5.1 Schematic of the general lead algorithm program flow. The algorithm iterates from index 0 to N . The numerical contact portion of the algorithm has a length of N slabs.

5.2.3. Simple scattering model in contact

Many types of modern semiconductor devices, e.g. RTD or TFET have complicated structures in the emitter and collector regions due to electrostatic band bending or hetero-epitaxial confinement. For these devices, the usual coherent tunneling approaches are not sufficient to model the charge injection from quasi-bound states in the emitter. The treatments of scattering based on self-consistent Born approximation [20][22][23] is numerically too expensive to be used in modeling a long structure with extreme thermalization of the carriers. One successful formula to handle this problem was presented first to study the RTD structures [55] as shown in Fig. 5.2(a). In this approach, the device is divided into three regions: the left and right reservoirs and the central device region. The reservoir regions are considered to be in thermal equilibrium, where an equilibrium Green's function is solved. The central device region is non-equilibrium where NEGF is solved for current density. As shown in Fig. 5.2(b) a quasi-bound state is present and inelastic scattering is required to feed this state. This scattering mechanism is depicted by an energy dependent optical potential $i^*\sigma$ added to the onsite energy, in the reservoirs. The electron density in the reservoir is calculated from the equilibrium relation, with Fermi distribution of the corresponding contact.

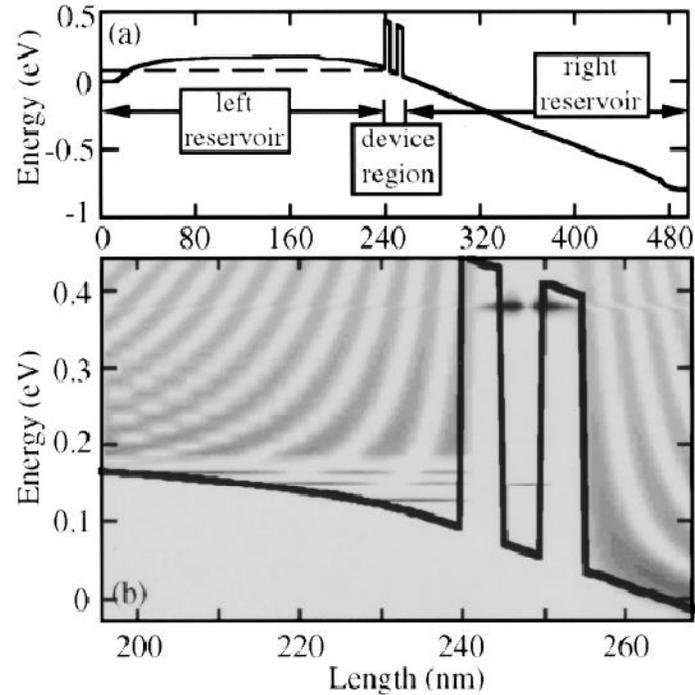


Fig. 5.2 (a) Conduction band profile for an $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}/\text{GaAs}$ RTD. The device is divided into three regions: left/right reservoirs and the central device. The reservoir regions are in thermal equilibrium. (b) The spectral function close to the central device region shows that the quasi-bound states are present and inelastic scattering is required to feed these states.

Since an optical potential $i^*\sigma$ is added to the reservoir, the same potential has to be added in the contact self-energy calculation in order to ensure Ohmic contact. This optical potential has to follow the band profile such that it vanishes within the bandgap to avoid unrealistic band tails. Since the bandedge changes with respect to electrostatic potential and confinement, it has to be determined during charge self-consistent calculation. Fig. 5.3 shows the spectrum for a 20nm long, 2.5nm thick Si UTB structure, for an in-plane momentum at 0.15 of the Brillouin zone (BZ). With $i^*\sigma$ in the bandgap the band tails are clearly seen close to the boundaries, while with $i^*\sigma$ only above the confined bandedge the unphysical band tails do not exist.

The $i^*\sigma$ represents the imaginary part of a retarded self-energy due to scattering in the reservoir. The relation between scattering rate and the scattering self-energy [23] is written as

$$S(k_r, k_{\parallel}, E) = -\frac{2}{\hbar} \text{Im} \int dr \exp(ik_r r) \Sigma^R(r, k_{\parallel}, E) \quad (5.2)$$

This can be simplified by setting the self-energy equal to $i^*\sigma$

$$S = \frac{2}{\hbar} \sigma \quad (5.3)$$

This defines a simple relation between the value of the optical potential and the scattering rate. Consequently, $i^*\sigma$ can be chosen wisely, i.e., according to experimentally measured scattering rate.

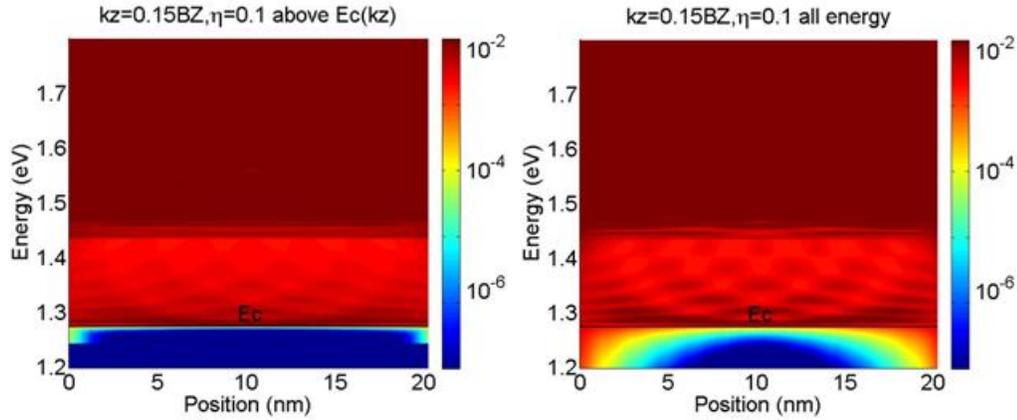


Fig. 5.3 Spectrum for Si UTB with in-plane momentum at 0.15BZ. Left: with $i^*\sigma$ only above the bandedge, the band tails do not exist; Right: with $i^*\sigma$ present in the bandgap, the band tails are clearly seen close to the boundaries.

5.3. Benchmark and Application

5.3.1. Si nanowire with periodic contact

The presented method is first verified on a $3\text{nm} \times 3\text{nm} \times 10\text{nm}$ homogeneous Si nanowire in the $sp^3d^5s^*$ TB representation [57]. Since the contact consists of pure Si and is considered to be periodic, the common boundary self-energy methods are applicable here and can serve as a verification baseline. The electronic transmission is solved with the NEGF method and the contact self-energies calculated by the transfer matrix method

[6] and the general lead method. In the general lead case, the numerical contact structure for the calculation includes 156 slabs in total. Fig. 5.4 shows that the transmission of the general lead case agrees well with the reference. All observed deviations are very close to the van-Hove singularities where convergence of the presented method is indeed the worst. These deviations are typically within a very small energy interval and become irrelevant in charge self-consistent calculations when the solution of the Poisson equation creates inhomogeneous potential profiles. In particular, for charge self-consistent I - V characteristics, such deviations have been found to be negligible. Shown in Fig. 5.5 is the I - V characteristics for the Si nanowire FET surrounded by a gate. It is an n - i - n structure with doping density 10^{20}cm^{-3} in the 8nm long gated region. It shows that the I - V for the two cases match very well also. This demonstrates that the general lead algorithm works well for ideal contacts. Table 5.1 shows the time and memory comparisons between the three contact self-energy methods: Sancho-Rubio, transfer matrix, and general lead. The transfer matrix case is solved based on the optimizations described in chapter 3.4. It is seen that the general lead method is the lowest in memory consumption, and is about 6x faster than the Sancho-Rubio method, even though it requires more iterations. This is because all the matrices in the general lead method is kept in the minimal slab size, while the matrices in the Sancho-Rubio method are with the contact Hamiltonian size.

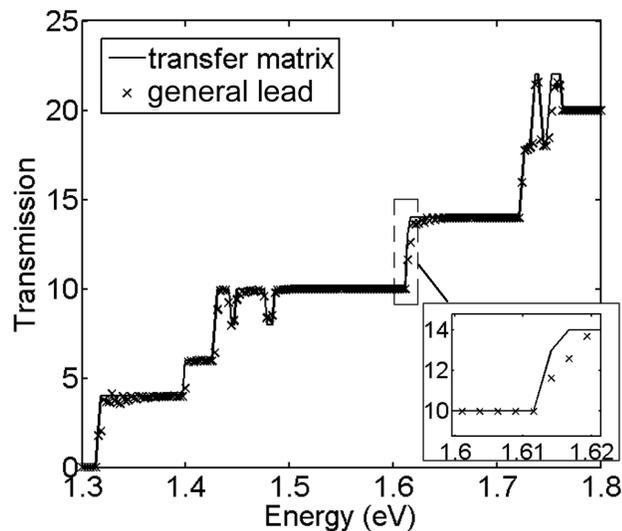


Fig. 5.4 Transmission comparison for a Si nanowire shows that the case with contact self-energy solved with general lead method (symbols) agrees well with the case with contact self-energy solved with transfer matrix method (solid line).

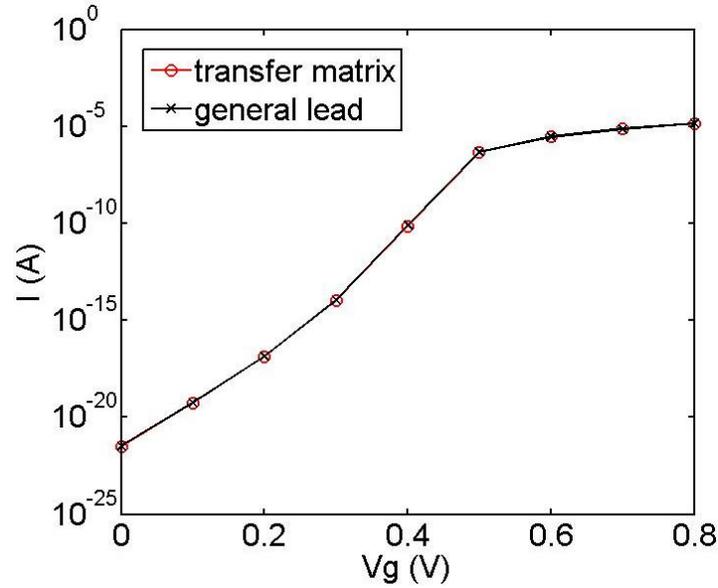


Fig. 5.5 I_d - V_g curve for a 20 nanometer long n - i - n Si nanowire FET with $V_{ds}=0.3$ V. The high doping density is 10^{20} cm^{-3} . The intrinsic gated region is 8nm long with oxide thickness 1nm. It shows that with the self-energy solved with general lead method (black line) the results agree well with the case when the self-energy is solved with transfer matrix method (red line).

Table 5.1 The measured time and memory comparisons between different contact self-energy methods: Sancho-Rubio, transfer matrix, and general lead. The calculations are done for the same Si nanowire as in Fig. 5.4, with 1 energy point only. The general lead method shows the lowest memory usage, and 6x faster than the Sancho-Rubio method.

	Transfer Matrix	Sancho-Rubio	General Lead
Time (s)	9.6	317	50.9
#iterations	NA	25	155
Memory (GB)	0.25	1.2	0.18

5.3.2. Graphene nanoribbon with trumpet shape contact

In contrast to the conventional contact methods, the presented general lead method allows solving contact self-energies of irregular shape contacts. This is demonstrated in the following with a 20nm long, 10nm thick graphene nanoribbon (GNR) illustrated in Fig. 5.6(a). Here, electrons are represented within the pd tight-binding model [58]. The contacts are considered as GNRs with a trumpet shape, i.e. the contact width increases constantly with increasing distance to the nanoribbon channel. Since neither the Sancho-

Rubio method nor the transfer matrix method can solve contacts of this kind, a benchmark with these methods is impossible. To verify the applicability of the general lead method in this contact structure, the transmission of this structure is solved in two ways, i.e. with two numerically different devices: In the case of device1, the active region where the NEGF equations are solved is limited to the central nanoribbon with a constant width of 10nm. In the case of device2, the active device extends 5nm beyond the central ribbon into the trumpet shaped contacts as shown in Fig. 5.6(a). In both cases, the contact self-energies describe the coupling to the remaining part of the trumpet shaped contacts, i.e. the contact/device interface has a different width in the two cases. Since the ballistic transmission describes electronic propagation from minus infinity to infinity along the transport direction, the transmission through the two devices, are physically equivalent. Indeed, the transmission results of the two cases are identical as shown in Fig. 5.6(b). This demonstrates the feasibility of applying the general lead method in non-trivial contact geometries.

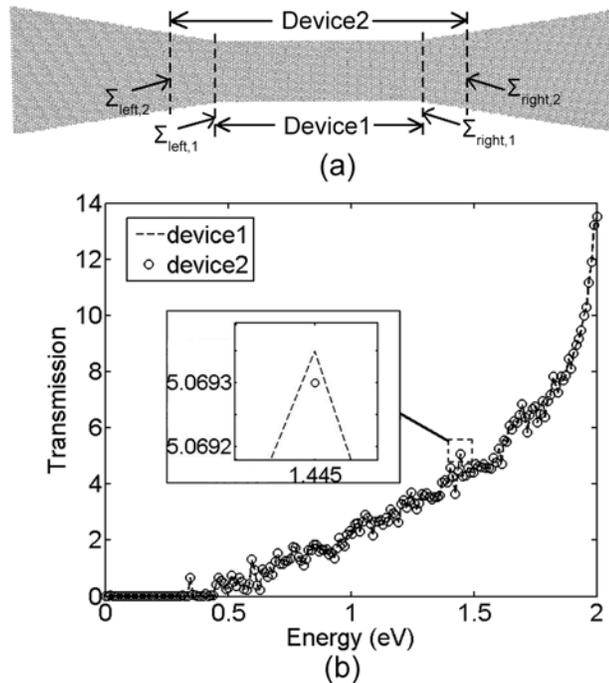


Fig. 5.6 (a) Graphene nanoribbon with trumpet shaped contacts: device 1 is defined only in the central square region; device 2 is the central region plus an extension of 5nm into the trumpet shape region. The contacts are the rest of the trumpet shape GNR. (b) The transmission for device 1 (dashed line) agrees well with device 2 (circle) showing that the general lead method works well in the trumpet shaped contacts.

5.3.3. Alloy disorder in contact

The third application of the presented approach addresses electron transport in explicitly disordered contacts and disordered devices. Most treatments of alloy disorder scattering in the NEGF literature involve a perturbative treatment [20]. The ability to model alloys in 3D atomistic representations affords the exploration of a non-perturbative alloy disorder scattering treatment. Supercell unfolding methods [59] showed that the typical virtual crystal approximation (VCA) cannot capture significant bandstructure effects such as band bowing. Approximate disordered nanowire bandstructures have been compared for AlGaAs wires [60] indicating the relationship between tunneling turn-on and the approximate bandstructure. These results also indicated that the atomistically disordered wire confines localized states which can result in resonant-like transport. In those simulations, however it cannot be concluded if the localization is induced by the centrally disordered device or from the attached homogenous contacts that sandwich the central device and form a heterostructure. With the approach described in this work one can treat quantum transport with disordered contacts.

Fig. 5.7(a) shows three nanowires with squared cross-sections of 3nm. The electrons are represented in $sp^3d^5s^*$ tight binding model [57]. The first device consists of pure Si; the second device of $\text{Si}_{0.5}\text{Ge}_{0.5}$ in the central 20nm only and ideal Si contacts surrounding it, and the third device consists of $\text{Si}_{0.5}\text{Ge}_{0.5}$ in both device and contacts. The Si and Ge atoms in all $\text{Si}_{0.5}\text{Ge}_{0.5}$ areas are randomly distributed within the atomically resolved structure. The Keating strain model [61] is applied to relax the structure prior to electron transport simulations. The changes of Hamiltonian due to strain are considered following the model of Boykin *et al.*[62]. All results that are subject to random atom distributions are averaged over 50 samples.

The transmission coefficients for the alloy cases and their error bars are shown in Fig. 5.7(b) with device length 20nm and 6nm. The calculation performed on VCA is shown for comparison. Fig. 5.7(b) illustrates for the case of pure Si contacts that the transmission through a 6nm long device is significantly larger than that of a 20nm long device. In contrast, contacts with random alloy yield virtually device length independent transmission coefficients. This indicates the tunneling of pure Si contact electrons through the random alloy $\text{Si}_{0.5}\text{Ge}_{0.5}$ devices: the longer the device, the more pronounced

is the dispersion mismatch between alloyed devices and ideal contacts and the less likely is tunneling through the device. Similar contact/device reflections were reported when incoherent scattering was included in the active device, but ballistic contacts were attached [23]. In contrast, when contacts and device face the same effective random alloy the DOS between the contacts and the device match and contact electrons face less reflection. As a result, the transmission remains the same when the device length changes. The small reduction of the device transmission with alloyed contacts in the 20nm long device compared to 6nm long one in Fig. 5.7(b) is due to a more pronounced weak localization in longer randomly alloyed devices. These results demonstrate that the presented method works well for contacts with alloy randomness.

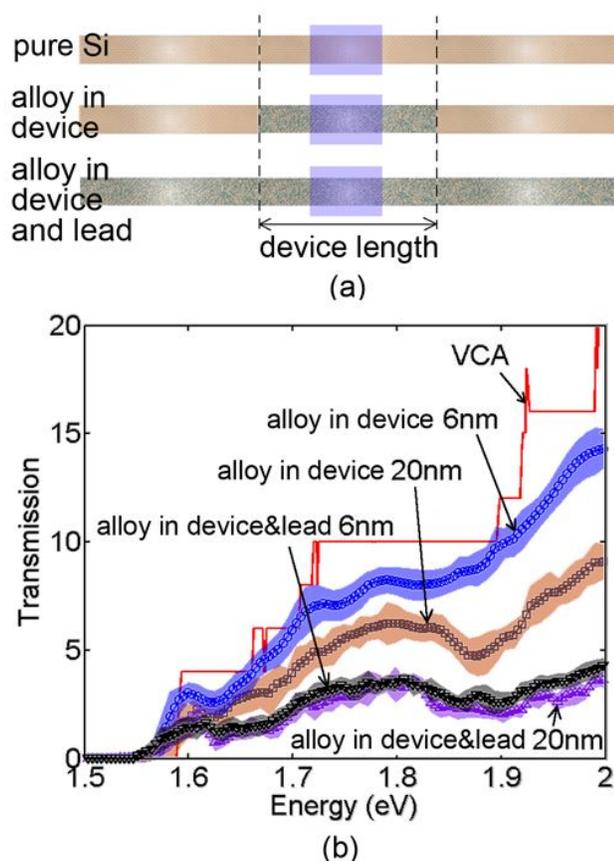


Fig. 5.7 (a) Schematic of randomly alloyed $\text{Si}_{0.5}\text{Ge}_{0.5}$ nanowires with pure Si contacts and contacts with $\text{Si}_{0.5}\text{Ge}_{0.5}$ random alloy. (b) Calculated transmissions for the alloy randomness structures as shown in (a), with device lengths of 20nm and 6nm, respectively. In the alloyed contact cases the transmissions show length independent properties suggesting that the DOS of the contacts matches well with the device. The transmission in VCA is shown for comparison.

Fig. 5.8 shows the I_d - V_g characteristics for the three cases shown in Fig. 5.7(a), with $V_{ds}=0.3V$. Their doping profile resembles an n - i - n structure with $10^{20}cm^{-3}$ in the n type doped regions. The total device length is 20nm with an oxide thickness of 1nm covering the 8nm long intrinsic region which is under the gate of the same length. As a result of the physics discussed in the previous paragraph, a big difference in the ON current density is observed between devices with pure Si contacts and devices with randomly alloyed $Si_{0.5}Ge_{0.5}$ contacts. With $V_g=0.5V$ and periodic Si contacts the current density is $3 \cdot 10^{-6}$ A. In contrast with random alloys in the contacts, the average current density is $1.74 \cdot 10^{-6}$ A. For comparison, the same calculation is performed within the VCA model which yields a current density of $3.14 \cdot 10^{-6}$ A. The random distribution of alloy atoms in the contacts reduces the ON-state current density by 42% compared to ideal contacts. This is also 45% lower than the VCA results. In the OFF-state, the differences are negligible.

The dramatic effect of the alloy disorder on electron flow through the 3nm thin wires supports some concerns some researchers harbor regarding the validity of the well-established empirical TB model. For 1D representation TB has proven to be very accurate and predictive. In a full 3D representation one may argue about a localization of the basis that is too tight resulting in excessively large scattering potentials. Yet this model has been connected well to previously known scattering potentials [63].

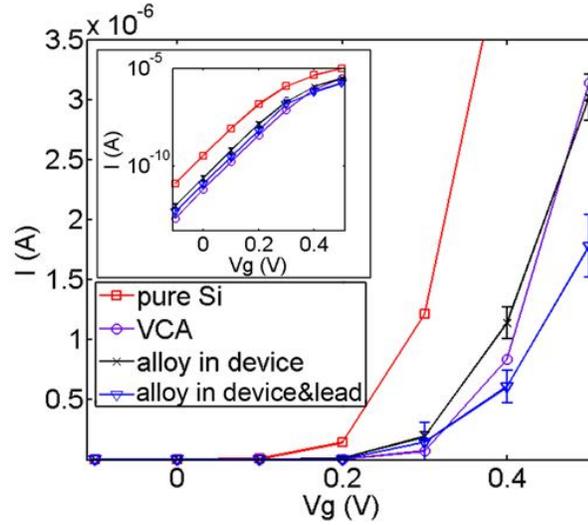


Fig. 5.8 I_d-V_g for the three devices shown in Fig. 5.7(a) and the VCA result. All current densities are solved with $V_{ds}=0.3\text{V}$. The random distribution of alloy atoms in the contacts reduces the ON-state current density by 42% compared to ideal contacts which suggests that having non-trivial contacts is critical in quantitative prediction of device performance.

5.4. Summary

A general lead method based on an extension of the CAP concept to arbitrary contact structures and disordered contacts is presented. The method adds a dephasing potential to the contact Hamiltonian to blur out any ballistic interference and solves the contact self-energy in an iterative way. The method is validated on regular Si nanowires and applied to trumpet shaped GNR contacts and contacts of randomly alloyed $\text{Si}_{0.5}\text{Ge}_{0.5}$. It is shown that alloy randomness in the contacts reduces the ON-state current in $3\text{nm}\times 3\text{nm}\times 20\text{nm}$ nanowire transistors by 45% compared to VCA calculations and by 42% compared to transistors with pure Si contacts, which suggests that the proper treatment of the contacts is critical in accurately predicting device performance.

6. SURFACE PASSIVATION METHOD

6.1. Introduction

Critical dimensions of modern semiconductor devices are in the domain of few thousands of atoms. Resolving the geometries and material compositions of these small devices in high detail is essential to accurately predict the electronic device performance. In particular, the surface treatment gets increasingly important since the surface-to-volume ratio increases with shrinking device dimensions. Theoretical device predictions require atomic resolutions of all device features and many band treatment of electrons as offered, e.g., by the empirical TB method [30][31]. The empirical TB method has been successfully applied to electronic band structure [18][32][33] and non-equilibrium transport calculations in modern nanodevices [19]-[21]. Surface atoms in TB contribute dangling bonds which often result in surface states within the material's band gap. This issue can be resolved when those dangling bonds are coupled to passivation atoms such as, e.g., hydrogen atoms [35]. The two common numerical passivation methods are to either explicitly include passivation atoms and their coupling to the surface atoms within the electronic Hamiltonian matrix [35], or to alter the orbital energies of the dangling bonds with a passivation potential [18]. The explicit inclusion of passivation atoms is a very general approach and applicable to any semiconductor surfaces. However, the rank of the Hamiltonian matrices can increase significantly with the explicit inclusion of passivation atoms and their orbital degrees of freedom. This increases the numerical load particularly for nanodevices with a high surface-to-volume ratio. For the case of zincblende and diamond crystal structures, Lee *et al.* have shown in [18] how to implicitly passivate sp^3 -hybridized dangling bonds with passivation potentials only. In this way, the passivation does not increase the rank of the Hamiltonian matrices and the numerical load stays the same. However, due to the assumed sp^3 -hybridization the model

of [18] considers only passivation of s and p orbitals. It is also restricted to sp^3 -hybridized bonding symmetries and does not distinguish between different passivation atoms (such as hydrogen and oxygen passivation). These aspects become increasingly relevant for state of the art nanodevices.

This chapter introduces a method to passivate dangling bonds in TB for arbitrary crystal structures and hybridization symmetries. This method distinguishes passivation atoms, since it uses *ab-initio* results for different passivation atoms as fitting targets. Similar to the method of Lee *et al.*, this method neither increases the rank of the electronic Hamiltonian nor the numerical complexity of solving band structure or electronic transport. In the following sections, the method is introduced and it is shown that it agrees with the one of Lee *et al.* for specific passivation parameters. The method is then applied to the passivation of Si (100) dangling bonds with SiO₂ in three different oxidation configurations. These different configurations are assessed with respect to their impact on the electronic properties and I - V characteristics of a UTB field effect transistor.

6.2. Method Details

6.2.1. Self-energy of the passivation atoms

The electronic Hamiltonian in the present method follows the standard TB approach for all non-surface atoms [34]. The Hamiltonian of each surface atom H_{SS} is setup as

$$H_{SS} = H_0 + \lambda_P I + \sum_{P=1}^{N_{db}} \Sigma_{SS,P} \quad (6.1)$$

Here, H_0 is the original Hamiltonian without passivation, N_{db} is the number of dangling bonds, λ_P is a surface potential, and $\Sigma_{SS,P}$ is a self-energy due to the coupling to the passivation atom P , which is given as

$$\Sigma_{SS,P} = H_{SP} (\varepsilon - H_P)^{-1} H_{PS} \quad (6.2)$$

Notice that (6.2) is inspired by the contact self-energies of the NEGF method [1]. The inversion in (6.2) represents the surface Green's function of the passivation atom P . It can be interpreted as a special case of the general lead method described in chapter 5,

since a surface Green's function is used to depict the passivation atoms and it is translated into a self-energy of the boundary of the device. In NEGF, ε represents the electronic energy. In this work, however, ε is a constant fitting parameter. H_P is the Hamiltonian of the passivation atom and H_{SP} is the coupling Hamiltonian between the surface atom and the passivation atom. If dangling bonds of two different surface atoms S and S' couple to the same passivation atom P , the passivation self-energy has interatomic contributions

$$\Sigma_{SS',P} = H_{SP}(\varepsilon - H_P)^{-1}H_{PS'} \quad (6.3)$$

Interatomic passivation self-energies beyond surface atoms that couple to the same passivation atom are ignored. All Hamiltonians H_0 , H_P , H_{SP} etc. are setup following the notation described in Table 2.1.

6.2.2. Parameterization

All required TB parameters, i.e. the onsite orbital energies of the passivation atom and the interatomic interactions are determined by fitting the TB band structures to the HSE06 exchange correlation functional [64] results of VASP [65]. In the VASP calculations, PAW pseudopotentials [66] for the electron-ion interaction are considered. Three top most valence bands and three lowest conduction bands are considered as the fitting targets. The energy window for fitting is set as 1.2eV around the middle of the band gap. The TB parameters of Si atoms in this work are taken from [36]. The present model is implemented and all TB results of this work are solved with the nanodevice simulator NEMO5 [37]. To compare TB results with optical band gaps of experiments, the exciton binding energy is estimated following [67] and subtracted from the calculated band gap. This estimation of the optical band gap was successfully applied to 2D MoS₂ in [67]. Ballistic transport in this work is solved with the QTBM method [6].

6.3. Benchmark and Application

6.3.1. Benchmark with known approach

The presented passivation method is validated against the known passivation method of [18]. It reproduces the Hamiltonian of [18] with the parameters $\varepsilon=1\text{eV}$, $V_{ss\sigma}=-2.739\text{eV}$, $V_{sp\sigma}=4.743\text{eV}$, $E_s=V_{sd\sigma}=V_{ss^*\sigma}=\lambda=0$. Given that only s -orbital parameters are needed to reproduce [18], one can interpret this known passivation method as a hydrogen passivation. Fig. 6.1 shows the benchmark on 3nm diameter Si and GaAs nanowires in 10 band TB model. From the band structures in Fig. 6.1 one can see the new method can exactly reproduce the published results with proper parameters.

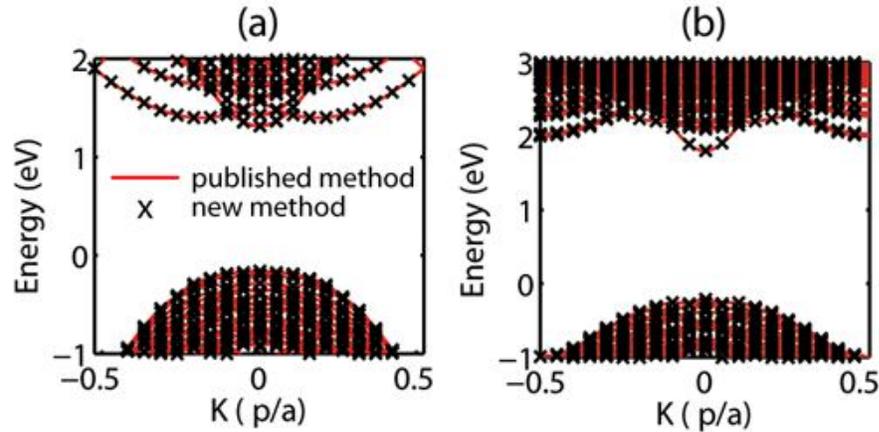


Fig. 6.1 Bandstructures for 3nm diameter nanowires with (a) Si and (b) GaAs, calculated with the published method (red line) and the new method (black symbol). The results show that the present method can reproduce the published method with proper parameters.

6.3.2. Application on Si/SiO₂ interface

In the following it is exemplified on a 2.2nm thick Si (100) quantum well structure embedded in SiO₂ that a careful treatment of the passivation atom type is needed to realistically predict device performances. It is shown in [67] that Si/SiO₂ has several interface configurations. The three configurations that differ most in their bandstructures are depicted in Fig. 6.2(a). In all configurations, the Si dangling bonds are partially

saturated with oxygen atoms (O1) of β -cristobalite SiO_2 . Remaining dangling bonds are either passivated with a double-bonded oxygen (O2) atom (DBM), or with two hydrogen atoms (HGM). In the bridging oxygen model (BOM), the dangling Si atoms that are not oxidized with SiO_2 are replaced with oxygen (O3) atoms. The coupling of dangling bonds with O3 is again modeled with the self-energy of (6.2) and (6.3). *Ab-initio* HSE06 calculations show that the quantum well bandstructures differ significantly for these three different oxidation configurations (Fig. 6.2 (b-d)). The *ab-initio* bandstructures in Fig. 6.2 are very well reproduced with $sp^3d^5s^*$ TB calculations of NEMO5 with the parameters of Table 6.1. It is worth to mention that the number of oxygen and hydrogen parameters in Table 6.1 is common for 10 band TB models [36]. The important fitting targets and their fit quality are listed in Table 6.2. The energy ε is for DBM $\varepsilon=0.008\text{eV}$, for BOM $\varepsilon=0.02335\text{eV}$, and for HGM $\varepsilon=-0.02324\text{eV}$.

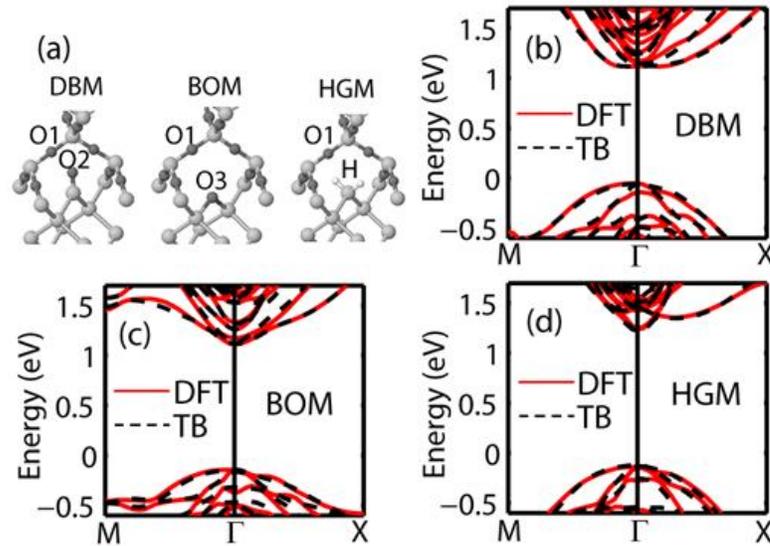


Fig. 6.2 (a) Atomic structure of the three Si/SiO₂ configurations described in the main text. Spheres represent Si (dark gray), oxygen atoms with different surroundings O1, O2, and O3 (black), and H (light gray). The dispersion relations of *ab-initio* (solid) and TB (dashed) calculations of a 2.2nm thick Si quantum well oxidized in (b) DBM, (c) BOM, and (d) HGM configurations agree well.

Table 6.1 Passivation parameters in units of eV.

	O1	O2	O3	H
E_s	-0.0232	0.000548	0.00586	-0.1232
E_p	6.0978	19.774	0.23768	NA
E_{s^*}	1.7973	0.75406	0.00756	
E_d	3.0662	5.0317	8.4569	
$V_{ss\sigma}$	-12.542	-0.20058	-0.1595	-7.8087
$V_{sp\sigma}$	0.00887	-0.208	0.23644	9.3511
$V_{sd\sigma}$	9.6611	-0.19985	-0.1992	0.076798
$V_{ss^*\sigma}$	1.8944	-0.31368	-0.00116	0.18137
$V_{ps^*\sigma}$	3.2026	0.17488	-0.07399	NA
$V_{pp\sigma}$	-0.1426	0.01789	0.62144	
$V_{pp\pi}$	0.2521	-8.3849	-0.36609	
$V_{pd\sigma}$	0.1188	0.71363	-1.7428	
$V_{pd\pi}$	0.9182	0.75347	0.19001	
$V_{dd\sigma}$	0.53801	9.26655	-6.1452	
$V_{dd\pi}$	1.3329	7.07982	-0.37154	
$V_{dd\delta}$	1.83785	8.510863	2.54124	
$V_{s^*s^*\sigma}$	1.64669	-2.96048	-0.99583	
$V_{s^*d\sigma}$	0.01886	-0.00109	0.12976	
λ	3.05833	5.430484	0.60796	

Table 6.2 Tight binding (TB) and *ab-initio* (DFT) band gaps E_g , valence E_v and conduction E_c band edges in eV, and effective masses for electrons m_e and holes m_h at the Γ point along X and M directions for the oxidation configurations of Fig. 6.2. The TB results match well with the DFT calculations.

	DBM		BOM		HGM	
	DFT	TB	DFT	TB	DFT	TB
E_v	-0.052	-0.060	-0.144	-0.138	-0.144	-0.115
E_c	1.12	1.115	1.111	1.121	1.237	1.22
E_g	1.172	1.175	1.255	1.259	1.381	1.335
m_{e_X}	1.701	1.644	0.410	0.388	0.193	0.183
m_{e_M}	1.941	1.927	0.665	0.672	0.2	0.226
m_{h_X}	-0.677	-0.665	-0.358	-0.312	-0.351	-0.341
m_{h_M}	-2.545	-2.497	-0.421	-0.408	-0.55	-0.518

The oxidation configurations DBM and BOM do not suppress surface states completely, but host significant electronic densities at the O2 and O3 atoms and Si atoms coupled to them. Fig. 6.3 shows the *ab-initio* projected wave functions of the valence band and the conduction band at Γ point, for the three oxidation configurations. This agrees with the findings of [69]. Such a surface density of states (DOS) introduces trapped states at the Si/SiO₂ interface which is expected to weaken the gate control of the transistor. Therefore, it is advisable to avoid these configurations in transistors. In contrast, the HGM configuration suppresses surface states very well, similar to the pure H atom passivation of [35].

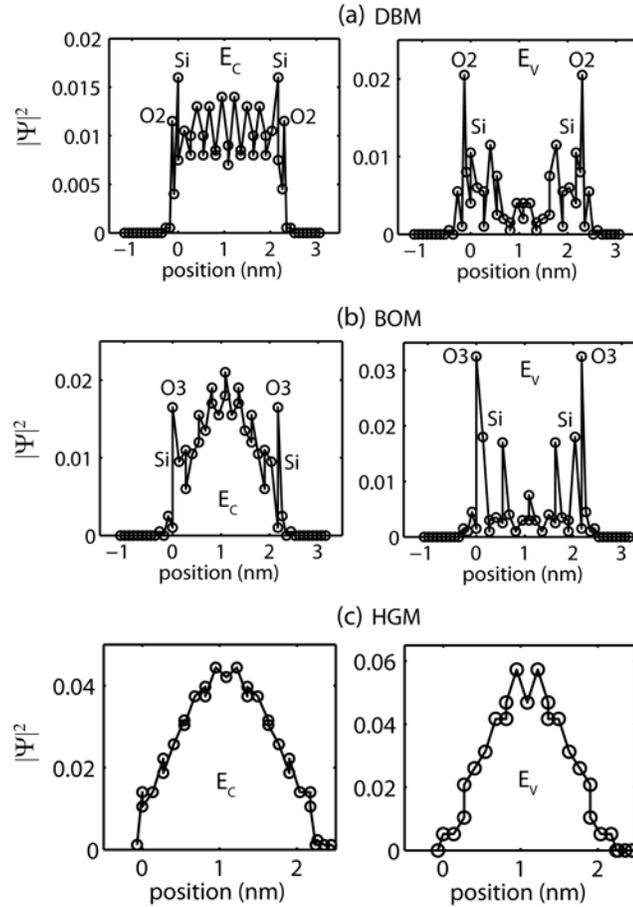


Fig. 6.3 *Ab-initio* wave functions for (a) DBM, (b) BOM, and (c) HGM oxidation configurations at Γ points. Left: for conduction band, right: for valence band. The results suggest that in the DBM and BOM configurations the surface states are not totally suppressed, while in the HGM case the surface states are well suppressed.

Since each O atom contributes eight orbitals (with six occupied) and each H atom only two (with one occupied), the device DOS should be the larger, the fewer H are used for passivation. Oxidation processes that only add O atoms (DBM) should have a higher DOS than cases that replace some Si atoms with O atoms (BOM). This is confirmed in Fig. 6.4 which shows the DOS in the 2.2nm Si quantum well passivated in different ways. Fig. 6.4(a) shows the results of pure H passivation following [18] and the HGM configuration of this work. The inset in Fig. 6.4(a) emphasizes the DOS of HGM is larger than the one of the pure H passivation for energies above the conduction band edge. Fig. 6.4 (b) shows the DOS of HGM is exceeded by BOM and even more by DBM results. It

can also be seen in Fig. 6.4 that the band gap of pure H passivation agrees with the HGM model, but DBM and BOM results deviate from that. This is elaborated in Fig. 6.5 which shows the thickness dependence of the Si quantum well optical band gaps solved in the TB oxidation models of this work. Experimental data of [70] are also shown for comparison. The calculated exciton binding energies for a 2.2nm Si quantum well in HGM and BOM configuration are 51meV and 75meV, respectively. This is of the same order as the exciton binding energy of homogeneous Si (20meV) [70]. For the DBM configuration, however, the exciton energy is 0.2eV due to its large effective masses (see Table 6.2). Tight binding calculations with the HGM model reproduce the measured optical band gaps of [70] very well, while the TB results in the DBM and BOM configurations are much lower. The small variation of DBM band gaps with the quantum well thickness in Fig. 6.5 agrees with *ab-initio* results of [68].

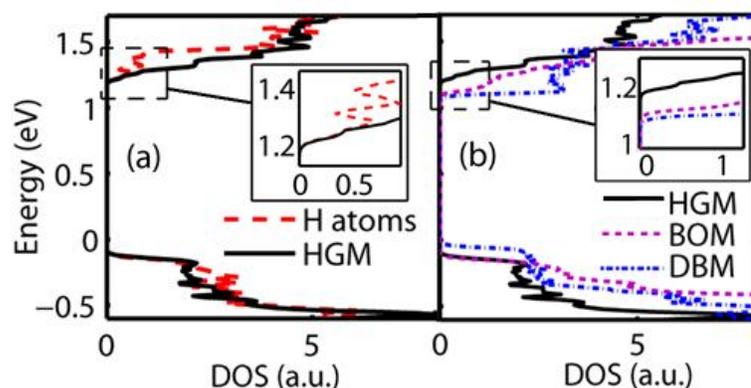


Fig. 6.4 Density of states comparisons of a 2.2nm Si quantum well for (a) HGM vs implicit H passivation, and (b) the three Si/SiO₂ configurations of Fig. 6.2. The results suggest that the DOS is higher with more O atom contributions in the system.

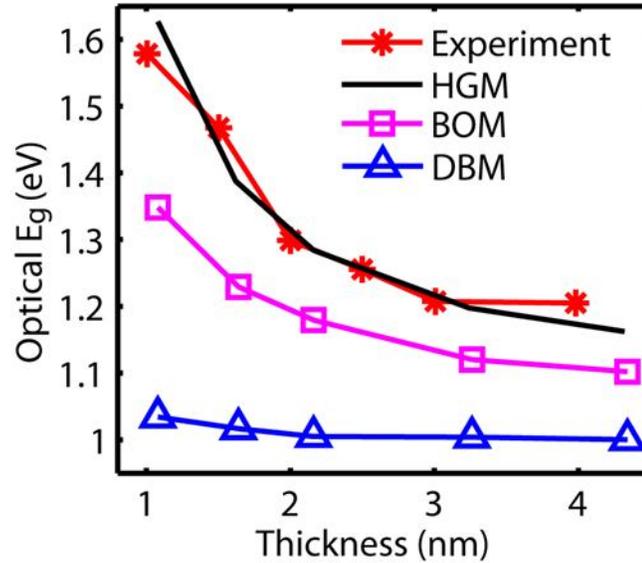


Fig. 6.5 Calculated optical band gaps of Si (100) quantum wells with varying thicknesses oxidized in the three configurations of Fig. 6.2. Experimental data of [70] (asterisks) are given for comparison. The HGM configuration produces optical band gaps in good agreement with experiment.

Fig. 6.6 shows the ballistic I_d - V_g characteristics of a Si UTB transistor with H atom passivation and oxidation with SiO_2 in the HGM configuration. The Si UTB of Fig. 6.6 follows the high performance logic technology requirements of ITRS 2020 [71]. The doping profile resembles an n - i - n UTB structure with $1.5 \cdot 10^{20} \text{cm}^{-3}$ in the n type doped regions. The thickness of the UTB is 3.3nm. The center 8.5nm long intrinsic channel is covered with a 2.8nm thick oxide layers on each UTB facet. Remaining UTB surfaces are passivated with H atoms following [18]. The threshold voltage V_{th} is defined at OFF current $I_{off}=10^{-10} \text{A/nm}$. The ON-state current is defined at $V_g - V_{th}=0.75 \text{V}$, and the source-drain bias $V_{ds}=0.75 \text{V}$. All these values agree with the ITRS 2020 requirements [71]. The subthreshold slope resulting from the HGM model is 147mV/dec, which exceeds the 121mV/dec predicted in the pure H atom passivation model of [18]. The ON/OFF ratio of the HGM model is $5.4 \cdot 10^3$ which is below the $7.7 \cdot 10^3$ of pure H atom passivation case. In conclusion, Si UTB transport calculations that model all dangling bonds passivated with only H atoms overestimate the transistor performance compared to calculations that consider gate areas covered with the best performing SiO_2 .

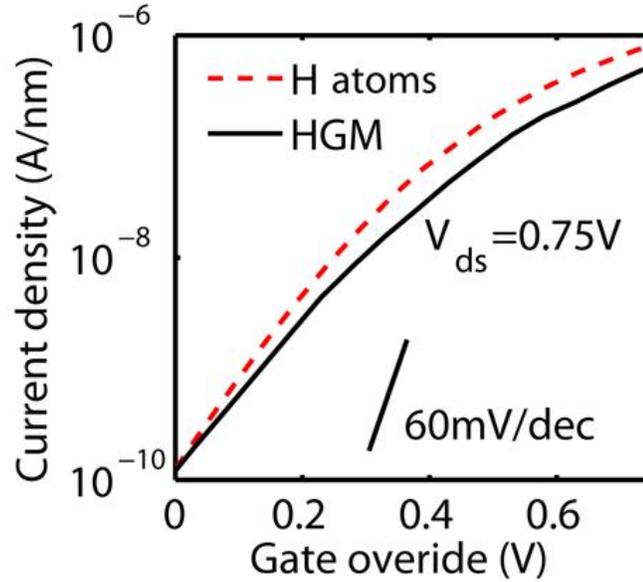


Fig. 6.6 I_d - V_g characteristic of the 3.3nm Si UTB transistor described in the main text when dangling bonds are passivated implicitly with H atoms following [18] (dashed) and with the present method for the HGM (solid) oxidization configuration. The results suggest that passivation with only H atoms overestimates the transistor performance.

6.4. Summary

In summary, this work introduces tight binding models for dangling bond passivation with SiO_2 in all relevant configurations. *Ab-initio* calculations served as input for fitting the passivation parameters. These models agree with an established H passivation model for a given parameter set. Tight binding band structure results of this work suggest that two of the SiO_2 configurations (DBM and BOM) have surface density of states, which should be avoided in transistors due to adverse impact on the performance. The HGM model shows bandgap dependency of thickness in good agreement with measurement and also well confined wave functions. It is also shown that passivation of all dangling bonds with only H atoms tends to overestimate the transistor performance.

7. BLACK PHOSPHORUS TRANSISTOR

7.1. Introduction

In past few years, extensive research efforts have been invested in the area of 2D materials such as graphene [72]-[75] and transition metal dichalcogenides (TMD) like MoS₂ [76]-[79]. For transistor applications, proper bandgaps and high carrier mobilities are the fundamental requirements. Graphene has excellent mobility but no bandgap, which is the major challenge for its application in logical devices [74]. On the other hand, materials in the TMD group such as MoS₂ usually have suitable intrinsic bandgaps thus gain increasing interest [76]. However, the obtained mobility of MoS₂ is ~300-400 cm²/Vs which is much lower than that of graphene [77]-[79].

In 2014, the first few-layer black phosphorus transistor was fabricated [80] which declares a new 2D material in the transistor community. Since then, significant progress has been made in both theoretical [81]-[86] and experimental [87]-[91] studies of black phosphorus, ranging from basic material properties to transport characteristics in transistors. Black phosphorus, like other 2D materials, has a puckered layer configuration, with van der Waals force to connect the different layers [82][83]. Unlike graphene, it has an intrinsic direct bandgap of ~0.3eV for its bulk configuration, which can open up to ~1.5eV in the monolayer configuration [87]. The bandgaps have layer dependent properties, with direct gaps for all different layers [82]. This feature makes it a good candidate for optical device applications. Furthermore, its carrier mobility is found to be >10⁴cm²/Vs both theoretically [83] and experimentally [92] which is beyond TMDs. Ambipolar transport characteristics are observed in transistor measurement, indicating the possibility to fabricate both *p*-type and *n*-type high performance transistors [89]. Measurements also show anisotropic transport properties along different in-plane directions [87], which can be traced back to the different electron and hole effective

masses along different crystal directions. This anisotropic behavior can lead to optimal design of transport along low effective mass direction, which produces high mobility, and with channel width along high effective mass direction which maintains high density of states [89].

Despite all the excellent features of black phosphorus, several challenges exist. Firstly, extracted mobility due to transport measurements in few-layer black phosphorus transistors is peaked at $\sim 300 \text{cm}^2/\text{Vs}$ [87], which is way below the $\sim 10^4 \text{cm}^2/\text{Vs}$ in magneto resistance measurement [92] and theoretical prediction [83], indicating strong scattering might exist. Secondly, surface roughness in black phosphorus samples is experimentally observed [90][91], but its impact in transport properties is not yet fully understood. These challenges point to the immediate requirement of comprehensive transport studies with proper multi-band representation. Existing theoretical works are mostly dispersion investigations [81]-[86]. Ballistic transport performed on monolayer black phosphorus transistors with four-band TB model qualitatively demonstrates the anisotropic properties [85], while quantitative comparison with measurement is not available. This work presents a quantum transport study of few-layer black phosphorus using NEGF [1] with ten-band TB model [6][36]. The TB parameters are determined using the *ab-initio* mapping method [35][36]. Acoustic phonon scattering [20][23] and surface roughness scattering [21] are considered in the transport simulation to understand the dominant scattering mechanism. Mobilities are extracted with given scattering mechanisms and compared with measurements. Afterwards black phosphorus transistors based on ITRS2020 logic device requirements [71] are simulated to give a performance prediction.

7.2. Method

7.2.1. TB parameterization

Empirical TB method has been successfully applied in nanoscale device simulations [19]-[21][31]-[36] given its great balance between accuracy and efficiency. For monolayer and bilayer black phosphorus, a four-band TB model was developed by Rudenko *et al.* in [84] and has been successfully applied to simulate ballistic transport in monolayer black phosphorus transistors [85]. However, as referred in [84], this four-band

model significantly underestimates the band splitting around the Γ valley of the valence band, which could be a potential problem in quantitative predictions of transport properties. Here, a ten-band TB model is proposed to give better description of the electronic band structures in few-layer and bulk black phosphorus.

The TB electronic Hamiltonian follows the standard $sp^3d^5s^*$ representation as shown in Table 2.1 for all the phosphorus atoms, except that different onsite energies are used for p_x and p_y orbitals to take into account the anisotropic properties. The 2nd nearest-neighbors (NN) of phosphorus atoms are considered, whose interatomic interaction parameters are altered following the Harrison's scaling law as described in [62]

$$U = U_0 \left(\frac{d_0}{d} \right)^\eta \quad (7.1)$$

Where d_0 is the 1st NN bond length between phosphorus atoms, U_0 is the 1st NN interatomic interaction, and η is a fitting parameter. Since the inter-layer atoms are connected with van der Waals forces which are different from in-plane atomic couplings, different parameters are used for in-plane (α) and inter-layer (β) interatomic interactions.

Three structures: monolayer, bilayer, and bulk black phosphorus are considered in the parameterization process. Fig. 7.1(a) and (b) show the top view and side view of a black phosphorus unit cell. Fig. 7.1(c) and (d) show the Brillouin zone (BZ) of monolayer and bulk black phosphorus. All required TB parameters are determined by fitting the TB band structures to the HSE06 exchange correlation functional [64] results of VASP [65]. In the VASP calculations, PAW pseudopotentials [66] for the electron-ion interaction are considered. Three top most valence bands and three lowest conduction bands are considered as the fitting targets.

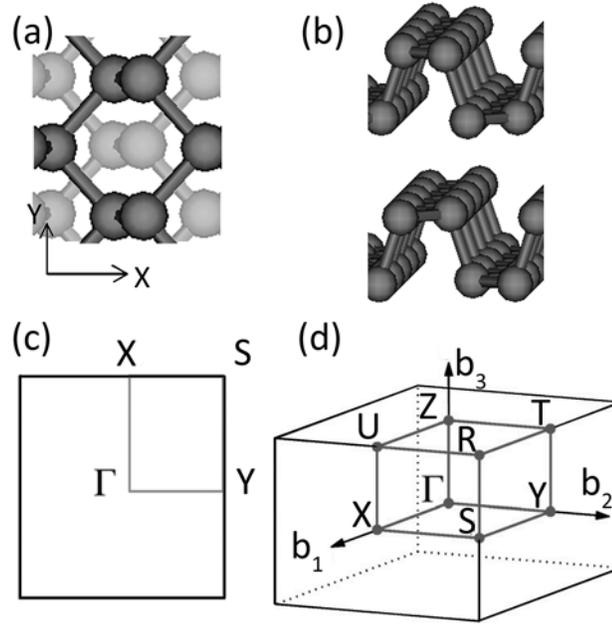


Fig. 7.1 Black phosphorus structure in (a) top view and (b) side view. The Brillouin zone for (c) monolayer and (d) bulk structures.

7.2.2. Acoustic phonon scattering

The scattering due to acoustic phonon is described by a scattering self-energy in the NEGF method, solved with self-consistent Born approximation [20]. Linear phonon dispersion is usually assumed so that the scattering potential can be written as

$$|U_q|^2 = \frac{\hbar D_{ac}^2}{2\rho v_s} q \quad (7.2)$$

Where D_{ac} is the deformation potential, ρ is the material density, and v_s is the sound velocity. For black phosphorus, density ρ is 2.69g/cm^3 [89]. Due to the anisotropic properties, sound velocities and deformation potentials differ along different directions. The sound velocities along armchair and zigzag directions (x and y in Fig. 7.1(a)) are 3.8km/s and 7.8km/s respectively [86]. Deformation potentials have both anisotropic and thickness dependent properties, which generally increase from monolayer to five-layers and then expect to saturate [83] for more layers. Low energy and high temperature condition is assumed here so that the scattering self-energy is written as

$$\Sigma(r_1, r_2, k, E) = \frac{D_{ac}^2 k_B T}{2\hbar \rho v_s^2} \delta(r_1, r_2) \int \frac{dk'}{2\pi} G(r_1, r_2, k', E) \quad (7.3)$$

Since few-layer black phosphorus assembles a UTB transistor, the k and k' in (7.3) represent the momentums perpendicular to the transport direction.

7.2.3. Surface roughness scattering

The surface roughness scattering is modeled by modifying the surface atoms in certain statistics and averaged over a couple of samples. The statistical properties of the surface are characterized by an autocovariance function [21][22]

$$ACVF_{ij} = rms^2 \exp(-\sqrt{2} |r_i - r_j| / L_c) \quad (7.4)$$

where rms is the root mean square of the surface height, L_c is the correlation length of the rough surface, and $|r_i - r_j|$ represents the distance between two phosphorus atoms at a surface. The roughness exists at both sides of the UTB facets, and the rough structure generation follows the same procedure as described in [21][22]. The implementation of the roughness generation is discussed in Appendix B. During the roughness generation, in the central gated region of the UTB, the phosphorus atoms to be removed will be replaced by oxide atoms, and vice versa. In the regions that are not gated, the phosphorus atoms to be removed will be deleted from the structure and the dangling bonds will be passivated by hydrogen atoms. The passivation follows the same method as described in chapter 6. For this work, $L_c=1.1\text{nm}$ and $rms=0.2\text{nm}$ is used for roughness in few-layer black phosphorus. In order to produce a reasonable rough surface, in the periodic direction of the UTB a structure wider than one unit cell is used. For monolayer to three-layer, a width of $\sim 6\text{nm}$ is used; for four-layer to six-layer, a width of $\sim 3\text{nm}$ is used; and for seven-layer to ten-layer, a width of $\sim 2\text{nm}$ is used. Surface roughness exists in both channel and the contact regions of NEGF calculation, whose self-energies are solved with the generalized lead method described in chapter 5 that has been demonstrated to work well for non-periodic contact structures.

7.2.4. Mobility calculation

In principle, the low field resistance of a channel length L can be written as

$$R(L) = R_0 + \frac{L}{n\mu e} \quad (7.5)$$

Where n is the charge density, e is the unit charge, and μ is the mobility. The first term R_0 presents the contact resistance and ballistic resistance, which is independent of channel length [93]. The diffusive resistance is due to given scattering mechanisms, which is length dependent. By solving the low field resistance for two different channel lengths L_1 and L_2 , the mobility can be calculated with

$$\mu = \left(\frac{L_1}{n_1} - \frac{L_2}{n_2} \right) / e(R_1 - R_2) \quad (7.6)$$

Where R_1 , R_2 are the corresponding resistances, and n_1 , n_2 are the charge densities, which are given from transport calculations. As demonstrated in [93], as long as the channel for transport calculation is sufficiently long, this method can provide an accurate estimation of mobility. In this work, the two black phosphorus channel lengths are chosen as 26nm and 30nm, with 6mV and 7mV applied voltages, respectively.

7.3. Results and Discussions

7.3.1. Band structure

Fig. 7.2 shows that the *ab-initio* band structures are well reproduced with $sp^3d^5s^*$ TB calculations with the parameters listed in Table 7.1. Unlike the four-band model [84], this ten-band model captures the valence band splitting at the Γ valley perfectly. For bilayer structure, the valence band splitting from *ab-initio* calculation is 0.906eV, and from TB result is 0.887eV. The important fitting targets and their fit qualities are listed in Table 7.2. A strong anisotropy of effective masses along x and y directions is noticed, especially for holes of monolayer structure, whose effective mass along y direction is two orders of magnitude larger than that along x direction. This observation is in good agreement with previous studies [82][83]. It is worthwhile to notice that the hole effective mass decreases

dramatically from monolayer to bilayer, which indicates that for few-layer black phosphorus hole mobility could be higher than that in monolayer and transport properties will be better.

Rough surface will introduce dangling bonds, which are required to be passivated. Fig. 7.3 (a) shows an example of surface roughness structure for a four-layer black phosphorus. The band structures calculated with and without passivation are shown in Fig. 7.3 (b). Hydrogen passivation is consider here with parameters $V_{ss\sigma}=-2.586\text{eV}$ and $V_{sp\sigma}=3.036\text{eV}$. With passivation the surface states are successfully removed.

Table 7.1 Tight binding parameters for black phosphorus in eV. In-plane parameters are denoted by α , and inter-layer parameters are denoted by β .

	α	β		α	β
E_s	-7.089	NA	E_{pz}	-3.218	NA
E_{px}	-2.841		E_{s^*}	19.967	
E_{py}	-2.125		E_d	6.799	
$V_{ss\sigma}$	-2.268	-5.34	$\eta_{ss\sigma}$	4.469	7.188
$V_{sp\sigma}$	3.851	15.153	$\eta_{sp\sigma}$	6.399	8.688
$V_{sd\sigma}$	-2.462	-0.065	$\eta_{sd\sigma}$	5.754	-1.972
$V_{ss^*\sigma}$	-1.314	-2.506	$\eta_{ss^*\sigma}$	4.679	3.924
$V_{ps^*\sigma}$	0.417	0.478	$\eta_{ps^*\sigma}$	-0.266	0.391
$V_{pp\sigma}$	4.945	3.902	$\eta_{pp\sigma}$	3.972	4.585
$V_{pp\pi}$	-1.489	-1.656	$\eta_{pp\pi}$	8.066	4.524
$V_{pd\sigma}$	-2.082	-2.566	$\eta_{pd\sigma}$	2.402	2.102
$V_{pd\pi}$	1.431	0.055	$\eta_{pd\pi}$	3.892	-0.969
$V_{dd\sigma}$	-1.355	0.786	$\eta_{dd\sigma}$	1.271	0.299
$V_{dd\pi}$	0.423	0.399	$\eta_{dd\pi}$	-0.769	-0.146
$V_{dd\delta}$	-0.675	-0.522	$\eta_{dd\delta}$	10.536	7.584
$V_{s^*s^*\sigma}$	-3.833	0.116	$\eta_{s^*s^*\sigma}$	2.963	0.437
$V_{s^*d\sigma}$	1.169	0.017	$\eta_{s^*d\sigma}$	28.313	-2.139

Table 7.2 Tight binding (TB) and *ab-initio* (DFT) band gaps E_g , valence E_v and conduction E_c band edges in eV, and effective masses for electrons m_e and holes m_h at the Γ point along X and Y directions for the monolayer, bilayer, and bulk black phosphorus.

The TB results match well with the DFT calculations.

	Monolayer		Bilayer		bulk	
	DFT	TB	DFT	TB	DFT	TB
E_v	-1.988	-1.989	-1.568	-1.571	-1.185	-1.188
E_c	-0.640	-0.616	-0.704	-0.728	-0.965	-0.963
E_g	1.348	1.373	0.864	0.843	0.220	0.226
m_{e_X}	0.187	0.183	0.186	0.185	0.174	0.153
m_{h_X}	0.171	0.157	0.166	0.160	0.140	0.135
m_{e_Y}	1.088	1.092	1.146	1.138	1.198	1.090
m_{h_Y}	10.08	8.87	1.709	1.424	0.681	0.661

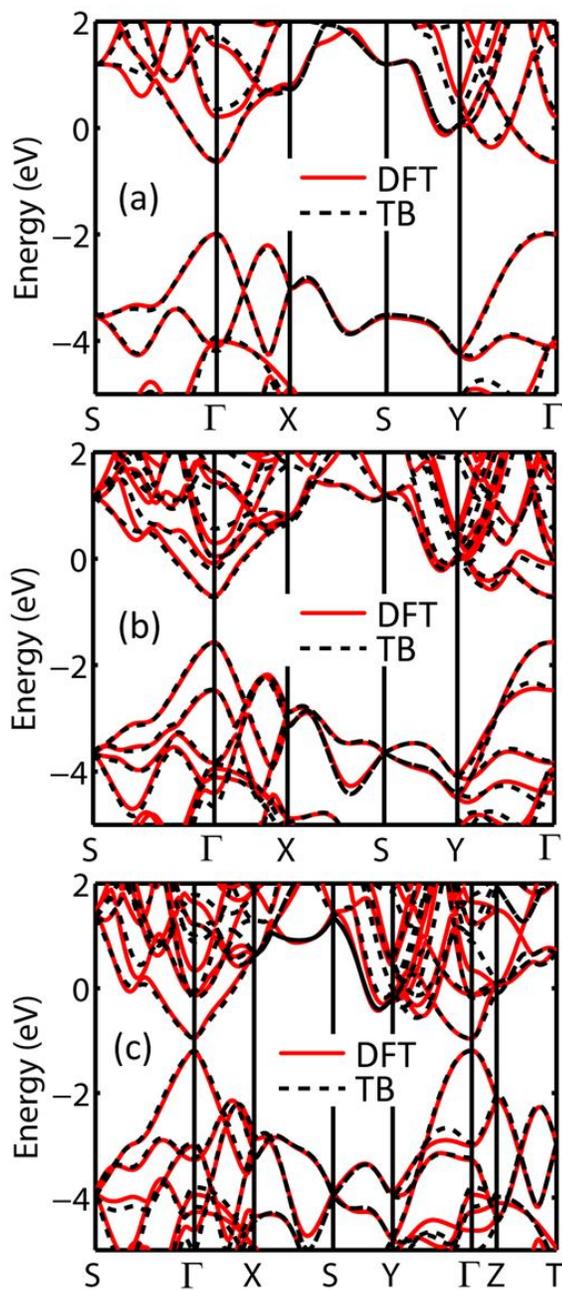


Fig. 7.2 Band structures for (a) monolayer, (b) bilayer, and (c) bulk black phosphorus solved with TB (black dashed line) agree well with *ab-initio* (red solid line) calculations.

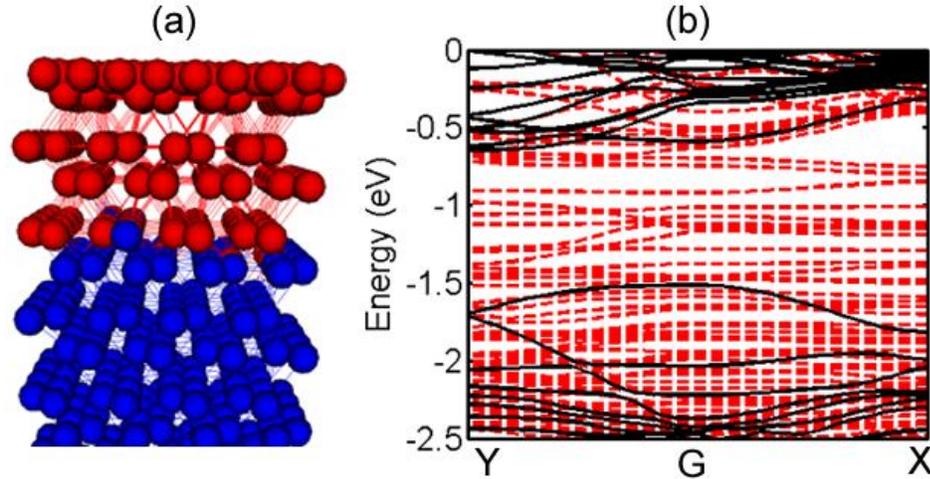


Fig. 7.3 (a) Interface roughness in a four-layer (1.8nm) black phosphorus, and (b) band structures with (black solid line) and without (red dashed line) passivation. It shows that with passivation the surface states are removed from the band gap.

7.3.2. Mobility

The calculated hole mobilities due to acoustic phonon scattering for different black phosphorus thicknesses are shown in Fig. 7.4(a). Anisotropy in the mobility is observed. The average mobility along armchair direction is $\sim 5 \cdot 10^3 \text{ cm}^2/\text{Vs}$, while the average mobility along zigzag direction is $\sim 2 \cdot 10^3 \text{ cm}^2/\text{Vs}$. This roughly agrees with the estimation in [83]. Along the armchair direction, the increase of mobility from monolayer to bilayer can be ascribed to comparably higher deformation potential in monolayer. From bilayer to four layers, the deformation potential increases gradually from 2.45eV to 3.16eV resulting in a gradual decrease of mobility. For thickness more than five layers, since the deformation potential saturates at 3.4eV, the mobility increases from $\sim 4.5 \cdot 10^3 \text{ cm}^2/\text{Vs}$ to $\sim 9.2 \cdot 10^3 \text{ cm}^2/\text{Vs}$ due to the decrease of hole effective mass. For the case along zigzag direction, it is interesting to notice that the mobility for monolayer is $\sim 4 \cdot 10^3 \text{ cm}^2/\text{Vs}$ which is twice of the average value of $\sim 2 \cdot 10^3 \text{ cm}^2/\text{Vs}$ for few-layer structures. This is attributed to the much smaller deformation potential ($\sim 0.15 \text{ eV}$) for monolayer, despite the fact that the hole effective mass is huge, which agrees with the findings in [83]. Unlike the armchair case, the mobility along zigzag direction oscillates with thickness, which can be ascribed to the combined effects of the changes in effective masses and the

deformation potentials for different number of layers. Nevertheless, the overall hole mobility due to acoustic phonon scattering is in the order of $\sim 10^3 \text{cm}^2/\text{Vs}$, which is much higher than the measured value of $\sim 300 \text{cm}^2/\text{Vs}$. This suggests that the acoustic phonon scattering is not the dominant part in the scattering mechanisms of black phosphorus transistors. Optical phonon scattering is also tested. The optical phonon energy for bulk [86] is used, and for the deformation potentials the same values as those in acoustic phonon are adopted [83]. The results show that the optical phonon scattering does not contribute much to the overall mobility with these artificial parameters.

Calculated hole mobilities due to surface roughness are shown in Fig. 7.4(b). The results are averaged over 50 samples. The anisotropic behavior is not observed from mobilities for thickness smaller than five-layer (2.3nm). For monolayer to three-layer, the mobility is $\sim 30 \text{cm}^2/\text{Vs}$ for both armchair and zigzag directions. For four-layer and five-layer the mobilities go up to $\sim 65 \text{cm}^2/\text{Vs}$ and $\sim 100 \text{cm}^2/\text{Vs}$, respectively. This indicates that the surface roughness breaks the anisotropic transport features in smaller number of layers. This can be explained as following: the roughness creates various irregular shapes of “islands” in the channel, whose dimension is related to *rms* and L_c . The influence of these “islands” is stronger for smaller number of layers due to the higher surface-to volume ratio, thus the anisotropic properties are destroyed. However, for larger number of layers the surface-to volume ratio is smaller thus the influence of the “islands” in the anisotropic transport properties becomes smaller. Consequently, for thicker structures the mobilities clearly show anisotropic properties. The mobilities for six-layer (2.8nm) are $\sim 150 \text{cm}^2/\text{Vs}$ and $\sim 100 \text{cm}^2/\text{Vs}$ along armchair and zigzag directions, respectively. For ten-layer (4.9nm) the mobilities are $\sim 440 \text{cm}^2/\text{Vs}$ and $\sim 240 \text{cm}^2/\text{Vs}$ along armchair and zigzag directions, respectively. The differences in mobility become larger for thicker structures, showing the anisotropic property is more pronounced for more layers. On the other hand, the mobility increases with respect to thickness for both armchair and zigzag directions, which can be attributed to the gradual decrease of hole effective masses along both directions. The measured mobilities [87] for several thicknesses are also shown in Fig. 7.4(b) for comparison. The calculated mobilities agree well with measurement, indicating that surface roughness might be the dominating scattering mechanism in black

phosphorus. The deviation between calculation and measurement in ten-layer structure can be ascribed to the geometry effect of the fabricated transistor in experiment. The experimental structure has metal contacts deposited on top of the channel, which introduces inter-layer resistance [87], whereas this effect is not considered in calculation. This resistance is more pronounced for thicker structures, leading to the deviations between calculation and measurement.

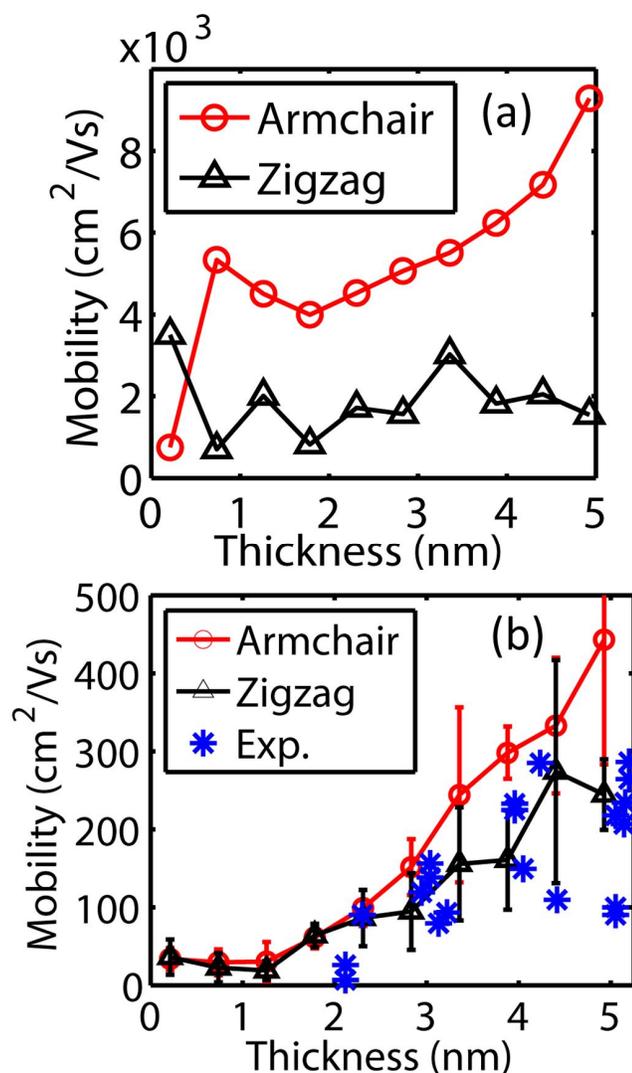


Fig. 7.4 Calculated hole mobility considering (a) acoustic phonon scattering and (b) roughness scattering for transport along armchair (red circle) and zigzag (black triangle) directions. Experimental data (blue star) from [87] is shown for comparison. The results show that the phonon assisted mobilities are much higher than measurement which suggests that acoustic phonon is not the dominating mechanism in the transport properties of black phosphorus. The mobilities of surface roughness agree well with measurement suggesting that roughness scattering might be the dominating mechanism.

7.3.3. ITRS 2020 target transistor

Fig. 7.5 shows the I_d-V_g characteristics of a black phosphorus UTB transistor with roughness scattering. Fig. 7.5 (a) shows an example of roughness in black phosphorus structure. The UTB follows the high performance logic technology requirements of ITRS 2020 [71]. The doping profile resembles a $p-i-p$ UTB structure with 10^{12}cm^{-2} in the p type doped regions. The thickness of the UTB is seven layers of black phosphorus (3.4nm). The center 8.5nm long intrinsic channel is covered with a 2.8nm thick oxide layers on both UTB facets. The threshold voltage V_{th} is defined at $I_{off}=10^{-10}\text{A/nm}$. The ON-state current is defined at $V_g-V_{th}=0.75\text{V}$, and the source-drain bias $V_{ds}=0.75\text{V}$. Roughness exists at both UTB facets. The results show obvious anisotropic transport properties. With transport along the armchair direction, the subthreshold slope is 76mV/dec, and ON/OFF ratio is ~ 1400 , with ON-state current $\sim 1.4\cdot 10^{-7}\text{A/nm}$. Transport along the zigzag direction, however, shows much weaker characteristics: the subthreshold slope is 82mV/dec, and ON/OFF ratio is ~ 130 , with ON-state current $\sim 1.3\cdot 10^{-8}\text{A/nm}$. However, the ON/OFF ratios in both cases are lower than ITRS2020 requirements of 10^4 .

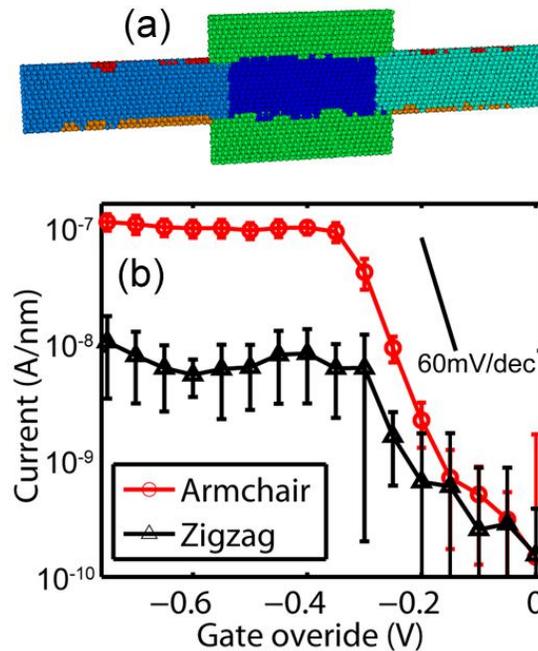


Fig. 7.5 (a) Structure with roughness (b) I_d-V_g characteristic for 3.4nm thick black phosphorus UTB transistors along armchair (red circle) and zigzag (black triangle) directions. The results show that the transport properties of the armchair direction outperform the zigzag direction, supporting the experimentally observed anisotropic transport properties.

7.4. Summary

In summary, this work presents transport study of few-layer black phosphorus transistors in empirical TB. A ten-band TB model is developed for black phosphorus and it reproduces the band structures perfectly compared to *ab-initio* results. Acoustic phonon scattering and surface roughness scattering in black phosphorus are investigated using NEGF and the mobilities are extracted. The results show that acoustic phonon scattering plays a little role in the overall mobility, while surface roughness produces mobilities in good agreement with measurement suggesting that it might be the dominating effect in the transport properties of black phosphorus. The I_d - V_g characteristics of black phosphorus UTB transistor under ITRS2020 requirement are simulated. The ON/OFF ratio is lower than the ITRS requirement. The results show clear anisotropic properties and suggest that the armchair direction outperforms the zigzag direction.

8. CONCLUSION AND OUTLOOK

In this thesis, several new methods were introduced to solve the existing challenges of quantum transport and its boundary conditions. Improvement of existing algorithms and their implementation details in NEMO5 were discussed in chapter 3, and a speed up of 5x was achieved in the self-energy calculations. In chapter 4, a low rank approximation method was introduced to reduce the computational effort in solving NEGF equations and the self-energies, and a significant speed up >100x is observed. Chapter 5 discussed a new general lead method which allows solving self-energies for non-ideal, non-periodic contacts. Chapter 6 introduced a new passivation model which allows passivation of arbitrary surfaces. These methods were applied to study the transport properties of few-layer black phosphorus transistors in chapter 7. Despite the promising results obtained, some issues should be further investigated for future improvements.

Firstly, the LRA method in TB requires the solution of an eigenvalue problem at each energy and momentum. This limits the speed up factor of the method. Although some efforts are taken in finding a common basis function for all the energies and momentums, it is not yet clear what the best way is to obtain this common basis. Furthermore, to solve transport with phonon scattering a real space basis is required. How to find a reduced real space basis in TB is not clear, since the correct transformation of the position operator in atomistic coordinate is not trivial. Also, a basis function that can couple with the RGF method will be promising, since that can take the advantage of the RGF method.

Further exploration of the eigenvalue solver is needed in the LRA method for the contact self-energy. In the transfer matrix method, one can obtain different spectrum of the contact modes by transforming the original matrix equations differently. The

eigenvalue problem discussed in chapter 4.4 is just one of the many possible ways. It is not yet clear which kind of spectrum will be the most efficient one to solve. Also, the best performing spectrum could be different in different eigenvalue solvers.

It has been shown in chapter 5.2.3 that the general lead method can adopt a simple scattering model in the contacts. It is worth to explore the coupling of the general lead method and the drift-diffusion approach in the contacts. Possible way to do this is to include the potential and chemical potential variations in the contacts, since the explicit contact structure is considered in the general lead method. In this way, one can generate finite current density in the contacts also, which can solve the vanishing current in the device/contact interface as shown in Fig. 4.1.

In chapter 6 a simple Si/SiO₂ interface model is studied with the new passivation method. The treatment of the oxide in the passivation so far is to include the oxygen atoms that are adjacent to the interface only. It is possible to include the full oxide structure and apply the general lead method to solve the surface Green's function of the oxide. In this way, it might be possible to include the effect of wave function penetration into oxide through a passivation treatment, rather than solving transport by explicitly including the oxide in the device Hamiltonian as in [21]. Also, in this interface model the structure is not fully relaxed. A more appropriate treatment is to fully relaxed the structure and include the effect of strain. To do so a proper strain model for the Si/SiO₂ interface is required. Furthermore, there are many different oxidation models for different Si surfaces. It is worth to explore those models and expand the passivation studies. Also, exploring the usage of this passivation model in the interfaces between high-k oxides and the different semiconductors would be very interesting.

The black phosphorus is a new 2D material recently joined the transistor community. The transport properties of black phosphorus are studied in this work to some extent. The roughness scattering seems to dominate the mobility, while acoustic phonon scattering seems not important. However, one should notice that in this work only acoustic phonon scattering is studied in this material, the effect of the optical branch is only tested with some artificial parameters. In measurement, one can observe temperature dependency of the mobility, which is clearly not roughness scattering related. Since acoustic phonon

scattering gives much higher mobilities than the measured values, it might be possible that optical phonon scattering is playing some roles in the overall mobility. It would be interesting to extract the proper parameters for the optical phonons of black phosphorus and investigate the impact of optical phonon scattering.

Also, tunneling FET studies based on TMD materials show that the effective masses and band gaps of different TMDs can lead to very different TFET performances [94]. It is suggested that 2D materials with low band gaps and low effective masses are good candidates for TFET designs. It would be interesting to apply few-layer black phosphorus structures in TFET since black phosphorus gives thickness dependent band gaps and low effective masses along armchair directions.

LIST OF REFERENCES

LIST OF REFERENCES

- [1] S. Datta, *Quantum Transport: Atom to Transistor*, Cambridge University Press, Cambridge, 2005.
- [2] J. Schwinger, “Brownian motion of a quantum oscillator”, *J. Math. Phys.*, Vol. 2, pp. 407, 1961.
- [3] L. P. Kadanoff and G. Baym, *Quantum statistical mechanics*, W. A. Benjamin, Inc, New York, 1962.
- [4] S. Fujita, “Partial self-energy parts of Kadanoff-Baym”, *Physica*, Vol. 30, pp. 848, 1964.
- [5] L. V. Keldysh, “Diagram technique for nonequilibrium processes”, *Sov. Phys. JETP*, Vol. 20, pp. 1018, 1965.
- [6] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, “Atomistic simulation of nanowire in the $sp^3d^5s^*$ tight-binding formalism: From boundary conditions to strain calculations”, *Phys. Rev. B*, Vol. 74, pp. 205323, 2006.
- [7] C. S. Lent and D. J. Kirkner, “The quantum transmitting boundary method”, *J. Appl. Phys.*, Vol. 67, pp. 6353, 1990.
- [8] T. Ando, “Quantum point contacts in magnetic fields”, *Phys. Rev. B*, Vol. 44, pp. 8017, 1991.
- [9] P. A. Khomyakov, G. Brocks, V. Karpan, M. Zwierzycki, and P. J. Kelly, “Conductance calculations for quantum wires and interfaces: Mode matching and Green’s functions”, *Phys. Rev. B*, Vol. 72, pp. 035450, 2005.
- [10] M. Sancho, J. Sancho, and J. Rubio, “Highly converge schemes for the calculation of bulk and surface Green functions”, *J. Phys. F: Met. Phys.*, Vol. 15, pp. 851, 1985.
- [11] N. Kharche, M. Luisier, T. B. Boykin, and G. Klimeck, “Electronic structure and transmission characteristics of SiGe nanowires”, *J. Comput. Electr.*, Vol. 7, pp. 350-354, 2008.
- [12] G. Klimeck, T. B. Boykin, M. Luisier, N. Kharche, and A. Schenk, “A study of alloyed nanowires from two perspectives: Approximate dispersion and transmission”, *AIP Conf. Proc.*, Vol. 893, pp. 711, 2007.

- [13] M. Auf der Maur, A. Pecchia, and A. Di Carlo, "Coupling drift-diffusion/NEGF for the simulation of InGaN/GaN LEDs", *2014 International Workshop on Computational Electronics*, pp. 95-96, 2014.
- [14] A. Paul, S. Mehrotra, M. Luisier, and G. Klimeck, "Performance prediction of ultrascaled SiGe/Si core/shell electron and hole nanowire MOSFETs", *IEEE Electron Dev. Lett.*, Vol. 31, pp. 278, 2010.
- [15] D. N. Nath, Z. C. Yang, C. Y. Lee, P. S. Park, Y. R. Wu, and S. Rajan, "Unipolar vertical transport in GaN/AlGaIn/GaN heterostructures", *Appl. Phys. Lett.*, Vol. 103, pp. 022102, 2013.
- [16] E. Jeon, H. Seo, C. W. Ahn, H. Seong, H. J. Choi, J. Kim, K. Kong, G. Buh, H. Chang, and J. Lee, "Resolving microscopic interfaces in Si_{1-x}Ge_x alloy nanowire devices", *Nanotechnol.*, Vol. 20, pp. 115708, 2009.
- [17] W. Park, J. Park, J. Jang, H. Lee, H. Jeong, K. Cho, S. Hong, and T. Lee, "Oxygen environmental and passivation effects on molybdenum disulfide field effect transistors", *Nanotechnol.*, Vol. 24, pp. 095202, 2013.
- [18] S. Lee, F. Oyafuso, P. von Allmen, and G. Klimeck, "Boundary conditions for the electronic structures of finite-extent embedded semiconductor nanostructures", *Phys. Rev. B*, Vol. 69, pp. 045316, 2004.
- [19] R. Lake, G. Klimeck, S. Datta, "Rate equations from the Keldysh formalism applied to the phonon peak in resonant-tunneling diodes", *Phys. Rev. B*, Vol. 47, pp. 6427, 1993.
- [20] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices", *J. Appl. Phys.*, Vol. 81, pp. 7845, 1997.
- [21] S. Kim, M. Luisier, A. Paul, T. B. Boykin, and G. Klimeck, "Full three dimensional quantum transport simulation of atomistic interface roughness in Silicon nanowire FETs", *IEEE Trans. Electr. Dev.*, Vol. 58, pp. 1371, 2011.
- [22] M. Luisier, *Quantum Transport Beyond the Effective Mass Approximation*, ph.D thesis, Swiss Federal Institute of Technology, Zurich, 2007.
- [23] T. Kubis, *Quantum Transport in Semiconductor Nanostructures*, ph.D thesis, Technische Universität München, Germany, 2009.

- [24] J. A. Driscoll, and K. Varga, "Calculation of self-energy matrices using complex absorbing potentials in electron transport calculations", *Phys. Rev. B*, Vol. 78, pp. 245118, 2008.
- [25] K. Varga, "Time-dependent density functional study of transport in molecular junctions", *Phys. Rev. B*, Vol. 83, pp. 195130, 2011.
- [26] L. Zhang, J. Chen, and J. Wang, "First-principles investigation of transient current in molecular devices by using complex absorbing potentials", *Phys. Rev. B*, Vol. 87, pp. 205401, 2013.
- [27] R. Wang, D. Hou, and X. Zheng, "Time-dependent density-functional theory for real-time electronic dynamics on material surfaces", *Phys. Rev. B*, Vol. 88, pp. 205126, 2013.
- [28] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically thin MoS₂: A new direct-gap semiconductor", *Phys. Rev. Lett.*, Vol. 105, pp. 136805, 2010.
- [29] L. Li, Y. Yu, G. J. Ye, Q. Ge, X. Qu, H. Wu, D. Feng, X. H. Chen, and Y. Zhang, "Black phosphorus field-effect transistors", *Nature Nanotechnol.*, Vol. 9, pp. 372, 2014.
- [30] J. Slater and G. Koster, "Simplified LCAO method for the periodic potential problem", *Phys. Rev.*, Vol. 94, pp. 1498, 1954.
- [31] J. Jancu, R. Scholz, F. Beltram, F. Bassani, "Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters", *Phys. Rev. B*, Vol. 57, pp. 6493, 1998.
- [32] G. Klimeck, F. Oyafuso, T. B. Boykin, C. R. Bowen, and P. V. Allmen, "Development of a nanoelectronic 3-D (NEMO 3-D) simulator for multi-million atom simulations and its application to alloyed quantum dots (invited)", *Computer Modeling in Engineering and Science (CMES)*, Vol. 3, pp. 601, 2002.
- [33] S. Goedecker and M. Teter, "Tight-binding electronic-structure calculations and tight-binding molecular dynamics with localized orbitals", *Phys. Rev. B*, Vol. 51, pp. 9455, 1995.
- [34] A. V. Podolskiy and P. Vogl, "Compact expression for the angular dependence of tight-binding Hamiltonian matrix elements", *Phys. Rev. B*, Vol. 69, pp. 233101, 2004.
- [35] Y. P. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, "Tight-binding analysis of Si and GaAs ultrathin bodies with subatomic wave-function resolution", *Phys. Rev. B*, Vol. 92, pp. 085301, 2015.

- [36] Y. P. Tan, M. Povolotskyi, T. Kubis, Y. He, Z. Jiang, G. Klimeck, and T. B. Boykin, “Empirical tight binding parameters for GaAs and MgO with explicit basis through DFT mapping”, *J. Comput. Electr.*, Vol. 12, pp. 56-60, 2013.
- [37] S. Steiger, M. Povolotskyi, H. H. Park, T. Kubis, and G. Klimeck, “NEMO5: a parallel multiscale nanoelectronics modeling tool”, *IEEE Trans. Nanotechnol.*, Vol. 10, pp. 1464, 2011.
- [38] MUMPS. URL <http://mumps.enseeiht.fr/>
- [39] LAPACK. URL <http://www.math.utah.edu/software/lapack/lapack-d/dgeev.html>
- [40] E. Polizzi and N. Ben Abdallah, “Self-consistent three-dimensional models for quantum ballistic transport in open systems”, *Phys. Rev. B*, Vol. 66, pp. 245301, 2002.
- [41] G. Mil’nikov, N. Mori, and Y. Kamakura, “Equivalent transport models in atomistic quantum wires”, *Phys. Rev. B*, Vol. 85, pp. 035317, 2012.
- [42] D. Mamaluy, M. Sabathil, and P. Vogl, “Efficient method for the calculation of ballistic quantum transport”, *J. Appl. Phys.*, Vol. 93, pp. 4628, 2003.
- [43] D. Mamaluy, D. Vasileska, M. Sabathil, T. Zibold, and P. Vogl, “Contact block reduction method for ballistic transport and carrier densities of open nanostructures”, *Phys. Rev. B*, Vol. 71, pp. 245321, 2005.
- [44] H. Ryu, *Electronic Structure and Quantum Transport in Controlled Impurity Systems*, ph. D thesis, Purdue University, United States, 2011.
- [45] P. Greck, *Efficient Calculation of Dissipative Quantum Transport Properties in Semiconductor Nanostructures*, ph.D thesis, Technische Universität München, Germany, 2012.
- [46] B. Haley, S. Lee, M. Luisier, H. Ryu, F. Saied, S. Clark, H. Bae, and G. Klimeck, “Advancing nanoelectronic device modeling through peta-scale computing and deployment on nanoHUB”, *J. Phys.: Conf. Ser.*, Vol. 180, pp. 012075, 2009.
- [47] M. Luisier and G. Klimeck, “Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering”, *Phys. Rev. B*, Vol. 80, pp. 155430, 2009.
- [48] Krylovschur in SLEPc library. URL slep.c.upv.es/documentation/reports/str7.pdf
- [49] PETSc/SLEPc team, private contact.

- [50]FEAST solver. URL <http://www.ecs.umass.edu/~polizzi/feast/>
- [51]CISS solver. URL http://zparecs.cs.tsukuba.ac.jp/?page_id=56
- [52]T. Schwamb, B. R. Burg, N. C. Schirmer, and D. Poulidakos, “On the effect of the electrical contact resistance in nanodevices”, *Appl. Phys. Lett.*, Vol. 92, pp. 243106, 2008.
- [53]F. Torricelli, E. C. P. Smits, J. P. Meijboom, A. K. Tripathi, G. H. Gelinck, L. Colalongo, Z. M. Kovacs-Vajna, D. M. de Leeuw, and E. Cantatore, “Transport physics and device modeling of zinc oxide and thin-film transistors-Part II: contact resistance in short channel devices”, *IEEE Trans. Electr. Dev.*, Vol. 58, pp. 3025, 2011.
- [54]W. S. Hwang, K. Tahy, X. Li, H. Xing, A. C. Seabaugh, C. Y. Sung, and D. Jena, “Transport properties of graphene nanoribbon transistors on chemical-vapor-deposition grown wafer-scale graphene”, *Appl. Phys. Lett.*, Vol. 100, pp. 203107, 2012.
- [55]G. Klimeck, R. Lake, R. C. Bowen, W. R. Frensley, and T. S. Moise, “Quantum device simulation with a generalized tunneling formula”, *Appl. Phys. Lett.*, Vol. 67, pp. 2539, 1995.
- [56]T. Kubis, C. Yeh, P. Vogl, A. Benz, G. Fasching, and G. Deutsch, “Theory of nonequilibrium quantum transport and energy dissipation in terahertz quantum cascade lasers”, *Phys. Rev. B*, Vol. 79, pp. 195323, 2009.
- [57]T. B. Boykin, G. Klimeck, and F. Oyafo, “Valence band effective-mass expressions in the $sp^3d^5s^*$ empirical tight-binding model applied to a Si and Ge parameterization”, *Phys. Rev. B*, Vol. 69, pp. 115201, 2004.
- [58]T. B. Boykin, M. Luisier, G. Klimeck, X. Jiang, N. Kharche, Y. Zhou, S. K. Nayak, “Accurate six-band nearest-neighbor tight-binding model for the π -bands of bulk graphene and graphene nanoribbons”, *J. Appl. Phys.*, Vol. 109, pp. 104304, 2011.
- [59]T. B. Boykin, N. Kharche, and G. Klimeck, “Brillouin-zone unfolding of perfect supercells having nonequivalent primitive cells illustrated with a Si/Ge tight-binding parameterization”, *Phys. Rev. B*, Vol. 76, pp. 035310, 2007.
- [60]T. B. Boykin, M. Luisier, A. Schenk, N. Kharche, G. Klimeck, “Atomistic, random-alloy calculations of the electronic structure and transmission characteristics of AlGaAs nanowires”, *IEEE Trans. Nanotechnol.*, Vol. 6, pp. 43-7, 2007.

- [61] P. N. Keating, "Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure", *Phys. Rev.*, Vol. 145, pp. 637, 1966.
- [62] T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, "Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory", *Phys. Rev. B*, Vol. 66, pp. 125207, 2002.
- [63] S. R. Mehrotra, A. Paul, and G. Klimeck, "Atomistic approach to alloy scattering in $\text{Si}_{1-x}\text{Ge}_x$ ", *Appl. Phys. Lett.*, Vol. 98, pp. 173503, 2011.
- [64] Y. S. Kim, K. Hummer, and G. Kresse, "Accurate band structures and effective masses for InP, InAs, and InSb using hybrid functional", *Phys. Rev. B*, Vol. 80, pp. 035203, 2009.
- [65] G. Kresse, and J. Furthmuller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set", *Phys. Rev. B*, Vol. 54, pp. 11169, 1996.
- [66] G. Kresse, and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method", *Phys. Rev. B*, Vol. 59, pp. 1758, 1999.
- [67] T. Cheiwchanchamnangij, W. R. L. Lambrecht, "Quasiparticle band structure calculation of monolayer, bilayer, and bulk MoS_2 ", *Phys. Rev. B*, Vol. 85, pp. 205302, 2012.
- [68] K. Seino, J. M. Wagner, and F. Bechstedt, "Quasiparticle effect on electron confinement in Si/SiO₂ quantum-well structures", *Appl. Phys. Lett.*, Vol. 90, pp. 253109, 2007.
- [69] P. Carrier, G. Abramovici, L. J. Lewis, and M. W. C. Dharma-wardana, "Electronic and optical properties of Si/SiO₂ superlattices from first principles: role of interfaces", *Mat. Res. Soc. Symp. Proc.*, Vol. 677, pp. AA4.10.1, 2001.
- [70] J. M. Wagner, K. Seino, F. Bechstedt, A. Dymiaty, J. Mayer, R. Rolver, M. Forst, B. Berghoff, B. Spangenberg, and H. Kurz, "Electronic band gap of Si/SiO₂ quantum wells: comparison of ab initio calculations and photoluminescence measurements", *J. Vac. Sci. Technol. A*, Vol. 25, pp. 1500, 2007.
- [71] International Technology Roadmap for Semiconductors 2013, <http://www.itrs.net/>.
- [72] K. S. Novoselov, V. I. Fal'ko, L. Colombo, P. R. Gellert, M. G. Schwab, and K. Kim, "A roadmap for graphene," *Nature*, vol. 490, pp. 192, 2012.
- [73] A. H. Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene," *Rev. Modern Phys.*, vol. 81, pp. 109, 2009.

- [74] F. Schwierz, "Graphene transistors," *Nature Nanotechnol.*, vol. 5, pp.487, 2010.
- [75] S. Z. Butler, S. M. Hollen, L. Cao, Y. Cui, J. A. Gupta, H. R. Gutierrez, T. F. Heinz, S. S. Hong, J. Huang, A. F. Ismach, E. Johnston-Halperin, M. Kuno, V. V. Plashnitsa, R. D. Robinson, R. S. Ruoff, S. Salahuddin, J. Shan, L. Shi, M. G. Spencer, M. Terrones, W. Windl, and J. E. Goldberger, "Progress, challenges, and opportunities in two dimensional materials beyond graphene," *ACS Nano*, vol. 7, pp. 2898, 2013.
- [76] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically thin MoS₂: A new direct-gap semiconductor," *Phys. Rev. Lett.*, vol. 105, pp. 136805, 2010.
- [77] X. Li, J. T. Mullen, Z. Jin, K. M. Borysenko, M. B. Nardelli, and K. W. Kim, "Intrinsic electrical transport properties of monolayer silicene and MoS₂ from first principles," *Phys. Rev. B*, vol. 87, pp. 115418, 2013.
- [78] M. S. Fuhrer and J. Hone, "Measurement of mobility in dual-gated MoS₂ transistors," *Nature Nanotechnol.*, vol. 8, pp. 146, 2013.
- [79] L. Zeng, Z. Xin, S. Chen, G. Du, J. Kang, and X. Liu, "Remote phonon and impurity screening effect of substrate and gate dielectric on electron dynamics in single layer MoS₂," *Appl. Phys. Lett.*, vol. 103, pp. 113505, 2013.
- [80] L. Li, Y. Yu, G. J. Ye, Q. Ge, X. Qu, H. Wu, D. Feng, X. H. Chen, and Y. Zhang, "Black phosphorus field-effect transistors," *Nature Nanotechnol.*, vol. 9, pp. 372, 2014.
- [81] K. Gong, L. Zhang, W. Ji, and H. Guo, "Electrical contacts to monolayer black phosphorus: A first-principles investigation," *Phys. Rev. B*, vol. 90, pp. 125441, 2014.
- [82] Y. Cai, G. Zhang, and Y. Zhang, "Layer-dependent band alignment and work function of few-layer phosphorene," *Sci. Rep.*, vol. 4, pp. 6677, 2014.
- [83] J. Qiao, X. Kong, Z. Hu, F. Yang, and W. Ji, "High-mobility transport anisotropy and linear dichroism in few-layer black phosphorus," *Nature Commun.*, Vol. 5, pp. 4475, 2014.
- [84] A. N. Rudenko and M. I. Katsnelson, "Quasiparticle band structure and tight-binding model for single- and bilayer black phosphorus," *Phys. Rev. B*, vol. 89, pp. 201408, 2014.
- [85] F. Liu, Y. Wang, X. Liu, J. Wang, and H. Guo, "Ballistic transport in monolayer black phosphorus transistors," *IEEE Trans. Electr. Dev.*, vol. 61, pp. 3871, 2014.

- [86] Z. Zhu and D. Tomanek, "Semiconducting layered blue phosphorus: A computational study," *Phys. Rev. Lett.*, vol. 112, pp. 176802, 2014.
- [87] H. Liu, A. T. Neal, Z. Zhu, Z. Luo, X. Xu, D. Tomanek, and P. D. Ye, "Phosphorene: an unexplored 2D semiconductor with a high hole mobility," *ACS Nano.*, vol. 8, pp. 4033, 2014.
- [88] J. D. Wood, S. A. Wells, D. Jariwala, K. Chen, E. Cho, V. K. Sangwan, X. Liu, L. J. Lauhon, T. J. Marks, and M. C. Hersam, "Effective passivation of exfoliated black phosphorus transistors against ambient degradation," *Nano Lett.*, vol. 14, pp. 6964, 2014.
- [89] H. Liu, Y. Du, Y. Deng, and P. D. Ye, "Semiconducting black phosphorus: synthesis, transport properties and electronic applications," *Chem. Soc. Rev.*, vol. 44, pp. 2732, 2015.
- [90] S. P. Koenig, R. A. Doganov, H. Schmidt, A. H. Castro Neto, and B. Ozyilmaz, "Electrical field effect in ultrathin black phosphorus." *Appl. Phys. Lett.*, vol. 104, pp. 103106, 2014.
- [91] R. A. Doganov, E. C. T. O'Farrell, S. P. Koenig, Y. Yeo, A. Ziletti, A. Carvalho, D. K. Campbell, D. F. Coker, K. Watanabe, T. Taniguchi, A. H. Castro Neto, and B. Ozyilmaz, "Transport properties of pristine few-layer black phosphorus by van der Waals passivation in an inert atmosphere," *Nature Commun.*, vol. 6, pp. 6647, 2015.
- [92] Y. Akahama and S. Endo, "Electrical properties of black phosphorus single crystals," *J. Phys. Soc. Japan*, vol. 52, pp. 2148, 1983.
- [93] Y. Niquet, V. Nguyen, F. Triozon, I. Duchemin, O. Nier, and D. Rideau, "Quantum calculations of the carrier mobility in thin films: methodology, Matthiessen's rule and comparison with semi-classical approaches," *J. Appl. Phys.*, vol. 115, pp. 054512, 2014.
- [94] H. Hatikhameneh, Y. Tan, B. Novakovic, G. Klimeck, R. Rahman, and J. Appenzeller, "Tunnel field-effect transistors in 2-D transition metal dichalcogenide materials", *Exploratory Solid-State Computational Devices and Circuits, IEEE Journal on*, vol. 1, pp. 12, 2015.

APPENDICES

A. REVISIT THE LRA BASIS FOR EFFECTIVE MASS

As discussed in chapter 4, in the TB representation one has to use the eigenfunctions from the device Hamiltonian with open boundary condition as the LRA basis sets, while in the effective mass approximation, the eigenfunctions from device Hamiltonian with Neumann boundary condition are good enough if ballistic transport is considered. The reason is that in effective mass case the Neumann boundary condition is a good approximation of the open boundary condition.

Consider an effective mass Hamiltonian, whose onsite energy is ε and coupling term is t . With lattice spacing a and effective mass m^* , the contact self-energy can be written as

$$\Sigma = te^{ika} \quad (\text{A.1})$$

where k is the wave vector. With the band minima defined at 0eV, the relation between k and energy is

$$E = \frac{\hbar^2 k^2}{2m^*} \quad (\text{A.2})$$

Consider the Taylor expansion of (A.1) around $k=0$, one can write

$$\Sigma = te^{ika} = t\left(1 + ika - \frac{k^2 a^2}{2} + \dots\right) \quad (\text{A.3})$$

Since in LRA only energies around $E=0\text{eV}$ are considered, one can drop the higher order terms in (A.3). Therefore, the Hamiltonian with open boundary condition (the matrix element at the device/contact interface) is written as

$$\varepsilon + \Sigma = \varepsilon + te^{ika} \approx \varepsilon + t \quad (\text{A.4})$$

This agrees with the Hamiltonian plus the Neumann boundary condition.

B. ROUGHNESS IMPLEMENTATION

The details of the surface roughness algorithm are described in [22], while the implementation is not clearly presented. Here, the implementation of the algorithm in NEMO5 is introduced. The basic equation for autocovariance function is given by (7.4). Its matrix form looks like

$$ACVF = \begin{bmatrix} A_{00} & A_{01} & A_{02} & \dots & \dots \\ & A_{11} & A_{12} & & \\ & & A_{22} & & \\ & & & \dots & \\ & & & & A_{NN} \end{bmatrix} \quad (\text{A.5})$$

Since it is a symmetry matrix containing the distances between surface atoms, only the upper triangular block is allocated. Solve the eigenvalue problem of $ACVF$

$$ACVF\varphi_i = \varepsilon_i\varphi_i \quad i = 1, 2, \dots, N \quad (\text{A.6})$$

The eigenvalues are randomized

$$\sigma_i = \sqrt{|\varepsilon_i|} \cdot rand() \quad i = 1, 2, \dots, N \quad (\text{A.7})$$

The roughness factor is defined as

$$SR_i = \varphi_i \cdot \sigma_i \quad i = 1, 2, \dots, N \quad (\text{A.8})$$

It is important to notice that the above equation relies on the explicit phase of an eigenvector φ_i , which is not deterministic. One way to solve this issue is to add a further constraint to φ_i , namely, ensure that its first element has a positive sign. Afterwards, loop

over all the atoms, and check their distances to the surface. If the distance of an atom is within a certain threshold, it writes

$$SR_j = \sum_{i=1, i \neq j}^N SR_i / |x_i - x_j| \quad (\text{A.9})$$

If $SR_i < 0$, the atom will be removed from the surface (no oxide exists), or the atom will be considered as part of the oxide (if oxide exists). If $SR_i > 0$ and an oxide exists, the oxide atom will be added to the surface.

Fig. A.1 shows several examples of surface roughness in a 20nm×20nm Si surfaces, with different correlation lengths. It is noticed that the roughness patterns are very different in these three cases. For $L_c=0.2\text{nm}$, the dimension of the rough “island” is pretty small, while for $L_c=1.4\text{nm}$ case, the “island” size is much larger. This is a simple demonstration of the roughness generation.

Fig. A.2 (a) shows the I - V for a 3nm diameter, 20nm long Si nanowire transistor, with and without roughness. It can be seen the roughness scattering lowers the current more, when the gate override is higher. This can be seen in Fig. A.2 (b). This effect agrees with previous studies in literature [22].

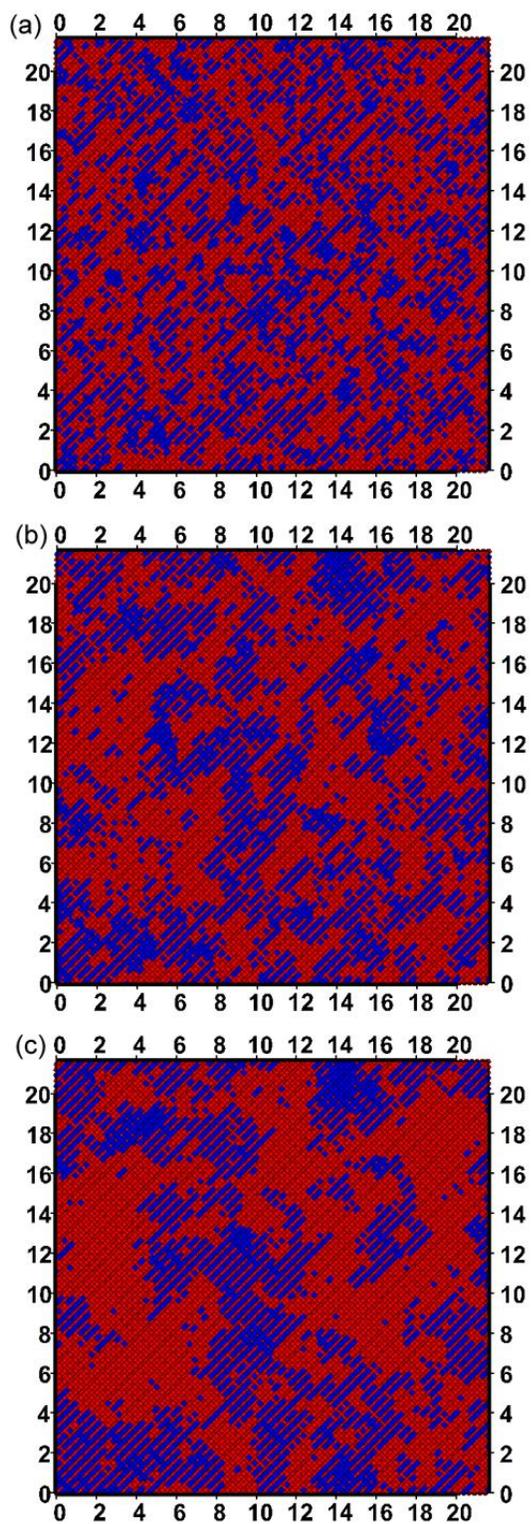


Fig. A.1 Surface roughness with (a) $L_c=0.2\text{nm}$, (b) $L_c=0.7\text{nm}$, and (c) $L_c=1.4\text{nm}$ show the right trend where larger correlation length produces larger rough “islands”.

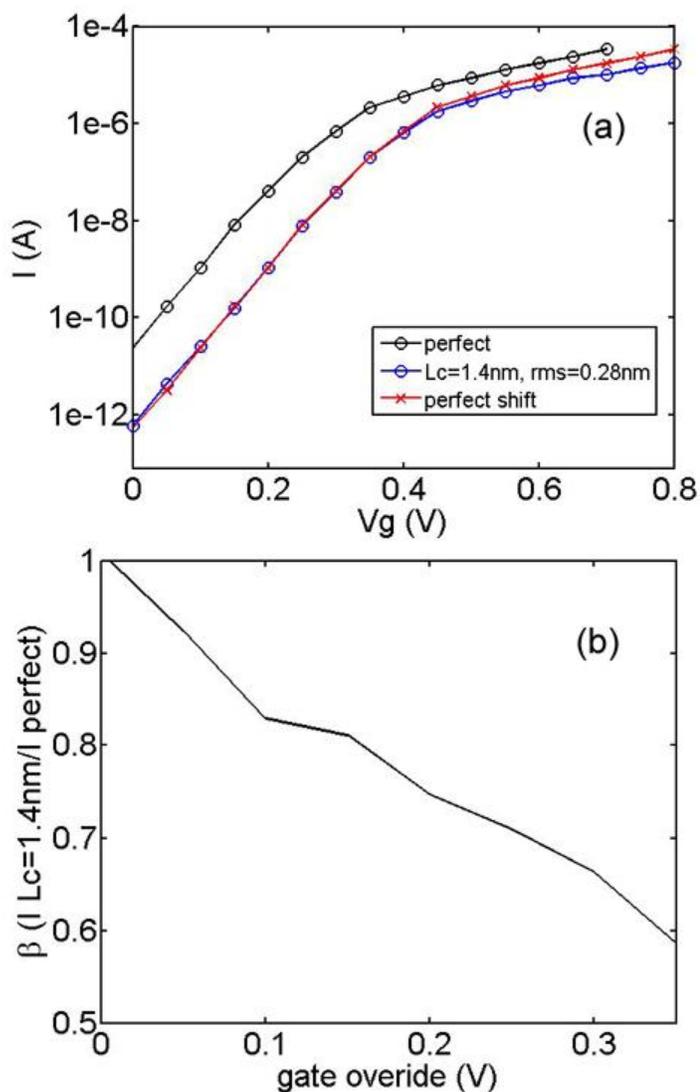


Fig. A.2 (a) I - V characteristic for a 3nm diameter, 20nm long Si nanowire in ballistic and with roughness scattering (b) the current density ratio between rough case and ballistic case. It shows that roughness scattering reduces the ON-state current density by 40% compared to ballistic calculations.

VITA

VITA

Yu He was born in Guangzhou, China, on December 17, 1984. From 2003 to 2007 he studied Microelectronics in Sun-Yat Sen University in China, and received his Bachelor's Degree in Engineering in 2007. In the same year he joined the Institute of Microelectronics in Tsinghua University, and received his Master's Degree in Electrical Engineering in 2010. In the same year he joined the department of Electrical and Computer Engineering, Purdue University, as a research assistant in the NEMO group. His research interest includes development of comprehensive software for nanoelectronic devices modeling, development of new modeling methods and algorithms, and physics of new materials and its application in nanoscale transistors.

PUBLICATIONS

PUBLICATIONS

- [1] Yu He, Yaohua Tan, James Charles, Gerhard Klimeck, and Tillmann Kubis, “Transport in few-layer black phosphorus with empirical tight binding”, under preparation.
- [2] Yu He, Yaohua Tan, Zhengping Jiang, Michael Povolotskyi, Gerhard Klimeck, and Tillmann Kubis, “Surface passivation in empirical tight binding”, submitted to *IEEE Trans. Electr. Dev.*, 2015.
- [3] Zhengping Jiang, Yeqing Lu, Yaohua Tan, Yu He, Michael Povolotskyi, Tillmann Kubis, Alan C Seabaugh, Patrick Fay, and Gerhard Klimeck, “Quantum transport in AlGaSb/InAs TFETs with gate field in-line with tunneling direction”, *IEEE Trans. Electr. Dev.*, Vol. 62, pp. 2445, 2015.
- [4] Yu He, Yu Wang, Gerhard Klimeck, and Tillmann Kubis, “Nonequilibrium Green’s Function Method: Non-trivial and disordered leads”, *Appl. Phys. Lett.*, Vol. 105, pp. 213502, 2014.
- [5] Yu He, Tillmann Kubis, Michael Povolotskyi, Jim Fonseca, and Gerhard Klimeck, “Quantum transport in NEMO5: algorithm improvements and high performance implementation”, *2014 International Conference on Simulation of Semiconductor Processes and Devices*, Yokohama, Japan, 2014.
- [6] Lang Zeng, Yu He, Michael Povolotskyi, Xiaoyan Liu, Gerhard Klimeck, and Tillmann Kubis, “Low rank approximation method for efficient Green's function calculation of dissipative quantum transport “, *J. Appl. Phys.*, Vol. 113, pp. 213707, 2013.
- [7] Jean Michel D. Sellier, James E. Fonseca, Tillmann Kubis, Michael Povolotskyi, Yu He, Hesameddin Ilatikhameneh, Zhengping Jiang, SungGeun Kim, Daniel. F. Mejia, Parijat Sengupta, Yaohua Tan, Gerhard Klimeck, “NEMO5, a parallel, multiscale, multiphysics, nanoelectronics modeling tool”, *2012 International Conference on Simulation of Semiconductor Processes and Devices*, Denver, USA, 2012.
- [8] Yaohua Tan, Michael Povolotskyi, Tillmann Kubis, Yu He, Zhengping Jiang, Gerhard Klimeck, and Timothy B. Boykin, “Empirical tight binding parameters for GaAs and MgO with explicit basis through DFT mapping”, *J. Comput. Electr.*, Vol. 12, pp. 56, 2012.

**American Physical Society
License Details**

Oct 07, 2015

This is an Agreement between Yu He ("You") and American Physical Society ("Publisher"). It consists of your order details, the terms and conditions provided by American Physical Society, and the payment instructions.

License Number	3723451083363
License date	Oct 07, 2015
Licensed content publisher	American Physical Society
Licensed content publication	Physical Review B
Licensed content title	Tight-binding analysis of Si and GaAs ultrathin bodies with subatomic wave-function resolution
Licensed copyright line	©2015 American Physical Society
Licensed content author	Yaohua P. Tan et al.
Licensed content date	Aug 4, 2015
Volume number	92
Type of Use	Thesis/Dissertation
Requestor type	Student
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs /tables/figures	1
Portion description	Figure 3
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
The requesting person/organization is:	Yu He
Order reference number	None
Title of your thesis / dissertation	ADVANCED BOUNDARY CONDITION METHOD IN QUANTUM TRANSPORT AND ITS APPLICATION IN NANODEVICES
Expected completion date	Dec 2015
Expected size (number of pages)	150

Total

0.00 USD

Terms and Conditions

Terms and Conditions

The American Physical Society (APS) is pleased to grant the Requestor of this license a non-exclusive, non-transferable permission, limited to [**print** and/or **electronic** format, depending on what they chose], provided all criteria outlined below are followed.

1. For electronic format permissions, Requestor agrees to provide a hyperlink from the reprinted APS material using the source material's DOI on the web page where the work appears. The hyperlink should use the standard DOI resolution URL, [http://dx.doi.org /{DOI}](http://dx.doi.org/{DOI}). The hyperlink may be embedded in the copyright credit line.
2. For print format permissions, Requestor agrees to print the required copyright credit line on the first page where the material appears: "Reprinted (abstract/excerpt/figure) with permission from [(FULL REFERENCE CITATION) as follows: Author's Names, APS Journal Title, Volume Number, Page Number and Year of Publication.] Copyright (YEAR) by the American Physical Society."
3. Permission granted in this license is for a one-time use and does not include permission for any future editions, updates, databases, formats or other matters. Permission must be sought for any additional use.
4. Use of the material does not and must not imply any endorsement by APS.
5. Under no circumstance does APS purport or intend to grant permission to reuse materials to which it does not hold copyright. It is the requestors sole responsibility to ensure the licensed material is original to APS and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover or table does not indicate that it was reprinted by APS, with permission from another source.
6. The permission granted herein is personal to the Requestor for the use specified and is not transferable or assignable without express written permission of APS. This license may not be amended except in writing by APS.
7. You may not alter, edit or modify the material in any manner.
8. You may translate the materials only when translation rights have been granted.
9. You may not use the material for promotional, sales, advertising or marketing purposes.
10. The foregoing license shall not take effect unless and until APS or its agent, Copyright Clearance Center (CCC), receives payment in full in accordance with CCC Billing and Payment Terms and Conditions, which are incorporated herein by reference.
11. Should the terms of this license be violated at any time, APS or CCC may revoke the license with no refund to you and seek relief to the fullest extent of the laws of the USA. Official written notice will be made using the contact information provided with the permission request. Failure to receive such notice will not nullify revocation of the permission.
12. APS reserves all rights not specifically granted herein.
13. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

Other Terms and Conditions

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

**AIP PUBLISHING LLC LICENSE
TERMS AND CONDITIONS**

Oct 07, 2015

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3723460784647
Order Date	Oct 07, 2015
Publisher	AIP Publishing LLC
Publication	Journal of Applied Physics
Article Title	Low rank approximation method for efficient Green's function calculation of dissipative quantum transport
Author	Lang Zeng, Yu He, Michael Povolotskyi, et al.
Online Publication Date	Jun 7, 2013
Volume number	113
Issue number	21
Type of Use	Thesis/Dissertation
Requestor type	Author (original article)
Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	8
Title of your thesis / dissertation	ADVANCED BOUNDARY CONDITION METHOD IN QUANTUM TRANSPORT AND ITS APPLICATION IN NANODEVICES
Expected completion date	Dec 2015
Estimated size (number of pages)	150
Total	0.00 USD

Terms and Conditions

AIP Publishing LLC -- Terms and Conditions: Permissions Uses

AIP Publishing LLC ("AIPP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material: "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], AIP Publishing LLC." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIPP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIPP, with permission, from another source. Under no circumstances does AIPP, purport or intend to grant permission to reuse material to which it does not hold copyright.
3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIPP reserves all rights not specifically granted herein.

4. The foregoing license shall not take effect unless and until AIPP or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
5. AIPP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIPP, AIP Publishing LLC, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
6. AIPP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIPP, AIP Publishing LLC, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIPP. This license may not be amended except in a writing signed by the party to be charged.
8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

**AIP PUBLISHING LLC LICENSE
TERMS AND CONDITIONS**

Oct 07, 2015

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3723461061870
Order Date	Oct 07, 2015
Publisher	AIP Publishing LLC
Publication	Journal of Applied Physics
Article Title	Low rank approximation method for efficient Green's function calculation of dissipative quantum transport
Author	Lang Zeng, Yu He, Michael Povolotskyi, et al.
Online Publication Date	Jun 7, 2013
Volume number	113
Issue number	21
Type of Use	Thesis/Dissertation
Requestor type	Author (original article)
Format	Print and electronic
Portion	Excerpt (> 800 words)
Will you be translating?	No
Title of your thesis / dissertation	ADVANCED BOUNDARY CONDITION METHOD IN QUANTUM TRANSPORT AND ITS APPLICATION IN NANODEVICES
Expected completion date	Dec 2015
Estimated size (number of pages)	150
Total	0.00 USD

Terms and Conditions

AIP Publishing LLC -- Terms and Conditions: Permissions Uses

AIP Publishing LLC ("AIPP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material: "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], AIP Publishing LLC." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIPP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIPP, with permission, from another source. Under no circumstances does AIPP, purport or intend to grant permission to reuse material to which it does not hold copyright.
3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIPP reserves all rights not specifically granted herein.

4. The foregoing license shall not take effect unless and until AIPP or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
5. AIPP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIPP, AIP Publishing LLC, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
6. AIPP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIPP, AIP Publishing LLC, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIPP. This license may not be amended except in a writing signed by the party to be charged.
8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

**AIP PUBLISHING LLC LICENSE
TERMS AND CONDITIONS**

Oct 07, 2015

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3723461262989
Order Date	Oct 07, 2015
Publisher	AIP Publishing LLC
Publication	Applied Physics Letters
Article Title	Quantum device simulation with a generalized tunneling formula
Author	Gerhard Klimeck,Roger Lake,R. Chris Bowen, et al.
Online Publication Date	Oct 23, 1995
Volume number	67
Issue number	17
Type of Use	Thesis/Dissertation
Requestor type	Student
Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	1
Title of your thesis / dissertation	ADVANCED BOUNDARY CONDITION METHOD IN QUANTUM TRANSPORT AND ITS APPLICATION IN NANODEVICES
Expected completion date	Dec 2015
Estimated size (number of pages)	150
Total	0.00 USD

Terms and Conditions

AIP Publishing LLC -- Terms and Conditions: Permissions Uses

AIP Publishing LLC ("AIPP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], AIP Publishing LLC." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIPP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIPP, with permission, from another source. Under no circumstances does AIPP, purport or intend to grant permission to reuse material to which it does not hold copyright.
3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIPP reserves all rights not specifically granted herein.
4. The foregoing license shall not take effect unless and until AIPP or its agent, Copyright

- Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
5. AIPP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIPP, AIP Publishing LLC, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
 6. AIPP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIPP, AIP Publishing LLC, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
 7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIPP. This license may not be amended except in a writing signed by the party to be charged.
 8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

Questions? customer@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

**AIP PUBLISHING LLC LICENSE
TERMS AND CONDITIONS**

Oct 07, 2015

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3723461478407
Order Date	Oct 07, 2015
Publisher	AIP Publishing LLC
Publication	Applied Physics Letters
Article Title	Non-equilibrium Green's functions method: Non-trivial and disordered leads
Author	Yu He,Yu Wang,Gerhard Klimeck, et al.
Online Publication Date	Nov 24, 2014
Volume number	105
Issue number	21
Type of Use	Thesis/Dissertation
Requestor type	Author (original article)
Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	4
Title of your thesis / dissertation	ADVANCED BOUNDARY CONDITION METHOD IN QUANTUM TRANSPORT AND ITS APPLICATION IN NANODEVICES
Expected completion date	Dec 2015
Estimated size (number of pages)	150
Total	0.00 USD

Terms and Conditions

AIP Publishing LLC -- Terms and Conditions: Permissions Uses

AIP Publishing LLC ("AIPP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material: "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], AIP Publishing LLC." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIPP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIPP, with permission, from another source. Under no circumstances does AIPP, purport or intend to grant permission to reuse material to which it does not hold copyright.
3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIPP reserves all rights not specifically granted herein.

4. The foregoing license shall not take effect unless and until AIPP or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
5. AIPP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIPP, AIP Publishing LLC, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
6. AIPP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIPP, AIP Publishing LLC, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIPP. This license may not be amended except in a writing signed by the party to be charged.
8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

**AIP PUBLISHING LLC LICENSE
TERMS AND CONDITIONS**

Oct 07, 2015

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3723470065649
Order Date	Oct 07, 2015
Publisher	AIP Publishing LLC
Publication	Applied Physics Letters
Article Title	Non-equilibrium Green's functions method: Non-trivial and disordered leads
Author	Yu He,Yu Wang,Gerhard Klimeck, et al.
Online Publication Date	Nov 24, 2014
Volume number	105
Issue number	21
Type of Use	Thesis/Dissertation
Requestor type	Author (original article)
Format	Print and electronic
Portion	Excerpt (> 800 words)
Will you be translating?	No
Title of your thesis / dissertation	ADVANCED BOUNDARY CONDITION METHOD IN QUANTUM TRANSPORT AND ITS APPLICATION IN NANODEVICES
Expected completion date	Dec 2015
Estimated size (number of pages)	150
Total	0.00 USD

Terms and Conditions

AIP Publishing LLC -- Terms and Conditions: Permissions Uses

AIP Publishing LLC ("AIPP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material: "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], AIP Publishing LLC." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIPP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIPP, with permission, from another source. Under no circumstances does AIPP, purport or intend to grant permission to reuse material to which it does not hold copyright.
3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIPP reserves all rights not specifically granted herein.

4. The foregoing license shall not take effect unless and until AIPP or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
5. AIPP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIPP, AIP Publishing LLC, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
6. AIPP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIPP, AIP Publishing LLC, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIPP. This license may not be amended except in a writing signed by the party to be charged.
8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
